

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2004, 02:02:48 ; Search time 278 Seconds  
(without alignments)  
7374.060 Million cell updates/sec

Title: US-10-019-495-26  
Perfect score: 3694  
Sequence: 1 ctggcagggttcctagtgcg.....gtcgtataaaaaaaaaaaaaa 3694

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*

3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*

4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*

5: /cgn2\_6/ptodata/2/ina/PCFUS\_COMB.seq:\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description          |
|------------|-------|-------------|--------|-------|----------------------|
| 1          | 370.2 | 10.0        | 3113   | 2     | US-08-993-228-20     |
| 2          | 57.4  | 1.6         | 1140   | 3     | US-09-023-173-4      |
| 3          | 56.2  | 1.5         | 23673  | 4     | US-09-773-816-1      |
| 4          | 55.2  | 1.5         | 1836   | 4     | US-09-475-515-47     |
| 5          | 55.2  | 1.5         | 1944   | 4     | US-09-475-515-46     |
| 6          | 55.2  | 1.5         | 2025   | 4     | US-09-475-515-45     |
| 7          | 54.8  | 1.5         | 2539   | 3     | US-09-000-016-3      |
| 8          | 54.8  | 1.5         | 2539   | 4     | US-09-514-340-3      |
| 9          | 54.8  | 1.5         | 2809   | 3     | US-09-000-016-1      |
| 10         | 54.8  | 1.5         | 2809   | 4     | US-09-514-340-1      |
| 11         | 53.6  | 1.5         | 1836   | 4     | US-09-475-515-41     |
| 12         | 53.6  | 1.5         | 1836   | 4     | US-09-475-515-44     |
| 13         | 53.6  | 1.5         | 1944   | 4     | US-09-475-515-37     |
| 14         | 53.6  | 1.5         | 1944   | 4     | US-09-475-515-38     |
| 15         | 53.6  | 1.5         | 1944   | 4     | US-09-475-515-40     |
| 16         | 53.6  | 1.5         | 1944   | 4     | US-09-475-515-43     |
| 17         | 53.6  | 1.5         | 2025   | 4     | US-09-475-515-36     |
| 18         | 53.6  | 1.5         | 2025   | 4     | US-09-475-515-39     |
| 19         | 53.6  | 1.5         | 2025   | 4     | US-09-475-515-42     |
| 20         | 53.6  | 1.5         | 2358   | 4     | US-09-475-515-50     |
| 21         | 53.6  | 1.5         | 2466   | 4     | US-09-475-515-49     |
| 22         | 53.6  | 1.5         | 2547   | 4     | US-09-475-515-48     |
| 23         | 53.6  | 1.5         | 4608   | 4     | US-09-475-515-76     |
| 24         | 53.6  | 1.5         | 4689   | 4     | US-09-475-515-74     |
| 25         | 53.2  | 1.4         | 2205   | 4     | US-09-252-991A-11217 |
| 26         | 53.2  | 1.4         | 2454   | 4     | US-09-252-991A-11038 |
| 27         | 52.6  | 1.4         | 1308   | 4     | US-09-501-115-31     |

|      |      |     |        |   |                      |                    |
|------|------|-----|--------|---|----------------------|--------------------|
| 28   | 51.6 | 1.4 | 909    | 4 | US-09-252-991A-13568 | Sequence 13568, A  |
| C 29 | 51.6 | 1.4 | 3129   | 4 | US-09-252-991A-13873 | Sequence 13873, A  |
| C 30 | 51.6 | 1.4 | 4188   | 4 | US-09-252-991A-13774 | Sequence 13774, A  |
| 31   | 51.6 | 1.4 | 8211   | 4 | US-09-252-991A-13656 | Sequence 13656, A  |
| 32   | 50.4 | 1.4 | 1317   | 4 | US-09-252-991A-6463  | Sequence 6463, Ap  |
| 33   | 50.2 | 1.4 | 1416   | 4 | US-09-252-991A-2955  | Sequence 2955, Ap  |
| C 34 | 50.2 | 1.4 | 1512   | 4 | US-09-252-991A-2728  | Sequence 2728, Ap  |
| 35   | 50   | 1.4 | 7218   | 1 | US-08-232-463-14     | Sequence 14, Appl  |
| C 36 | 49.8 | 1.3 | 2435   | 4 | US-09-484-970B-134   | Sequence 134, Appl |
| 37   | 49.8 | 1.3 | 152331 | 3 | US-09-128-155-16     | Sequence 16, Appl  |
| 38   | 49.4 | 1.3 | 3177   | 1 | US-08-042-747A-4     | Sequence 4, Appli  |
| C 39 | 49   | 1.3 | 364    | 4 | US-09-621-976-17202  | Sequence 17202, A  |
| 40   | 48.8 | 1.3 | 408    | 4 | US-09-252-991A-3047  | Sequence 3047, Ap  |
| C 41 | 48.8 | 1.3 | 1752   | 4 | US-09-679-686B-17    | Sequence 17, Appl  |
| 42   | 48.8 | 1.3 | 13842  | 3 | US-09-105-537-30     | Sequence 30, Appl  |
| 43   | 48.8 | 1.3 | 36778  | 3 | US-09-105-537-5      | Sequence 5, Appli  |
| 44   | 48.8 | 1.3 | 38506  | 3 | US-09-320-878-19     | Sequence 19, Appl  |
| 45   | 48.8 | 1.3 | 38506  | 4 | US-09-141-908-1      | Sequence 1, Appli  |

ALIGNMENTS

RESULT 1

US-08-993-228-20

; Sequence 20, Application US/08993228

; Patent No. 5976838

; GENERAL INFORMATION:

; APPLICANT: Jacobs, Kenneth

; APPLICANT: McCoy, John M.

; APPLICANT: LaVallie, Edward R.

; APPLICANT: Racie, Lisa A.

; APPLICANT: Merberg, David

; APPLICANT: Treacy, Maurice

; APPLICANT: Spaulding, Vikki

; APPLICANT: Agostino, Michael J.

; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

; TITLE OF INVENTION: ENCODING THEM

; NUMBER OF SEQUENCES: 33

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.

; STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: MA

; COUNTRY: U.S.A.

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/993,228

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Sprunger, Suzanne A.

; REGISTRATION NUMBER: 41,323

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 498-8284

; TELEFAX: (617) 876-5851

; INFORMATION FOR SEQ ID NO: 20:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3113 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

US-08-993-228-20

Query Match 10.0%; Score 370.2; DB 2; Length 3113;  
Best Local Similarity 56.4%; Pred. No. 3.4e-71;  
Matches 874; Conservative 0; Mismatches 606; Indels 69; Gaps 7;



```
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: codon-optimized RAmY3D signal-prosubtilisin BPN'
US-09-023-173-4

Query Match      1.6%; Score 57.4; DB 3; Length 1140;
Best Local Similarity 43.4%; Pred. No. 0.0022;
Matches 262; Conservative 0; Mismatches 341; Indels 0; Gaps 0;

QY 238 AACGTGCGGGTGTACTGCGGCCCTCTGGTGGAGAGAGACCCACCATGAAGCTG 297
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
379 AACGTCAAGGTGCGCGTCAATCGACAGCGGCATCGACAGCAGCCACCCGACCTCAAGGTC 438
QY 298 TGGTGTCCCGCGGCGTCAACCTGAGCGGGTGGAGGCCCAATGAGGACGACGCTGGGAAT 357
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
439 GCCGGCGGAGTAGCATGGTCCCGAGCGAGACCAACCCGTTCCAGGACACCAACAGCCAT 498
QY 358 GGAGTCAAGCCAGCGCCAGGCGCGGATCCCTGACCTGCGACCGCGAAGGAGCGCGAG 417
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
499 GGCACCCACGTGCGCGGCACCGCTGCGGCCCTCACCACAGCATCGGCGTCTCGCGTC 558
QY 418 CCCAAGAGCGCCACACGCTCTCCGAGAGAGAGGCCAAGGAGCTCCCTGAAATGGAC 477
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
559 GCCCGAGCGCCAGCCCTCTACGCGCTCAAGGTACTCTGGCGCGACGGCAGCGCCAGTAC 618
QY 478 GCCACCTCCAGCGGGTGTGGATCCTGACCAAGCACCTGACCAACAGCAAGTGGTGATC 537
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
619 AGCTGGATCATCAACGGCATCGAGTGGGCCATCGCCCTCAAGCCCGCTCGACAAGGCCGTG 678
QY 538 ATCGAGCGCCACAGCGCGGCGGACCGGTGGTGGACCAAGTTCACCGTCTGCAACGGCAGCTG 597
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
679 AGCTCGGCGCGGAGCGCGGAGCGCGCCCTCAAGCCCGCTCGACAAGGCCGTGCGCC 738
QY 598 CTGTGATCTCCAGCATCCCCCGGCGGCGGACAGCGAGCACTACCTCCCGGGAGATGTTT 657
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
739 AGCGCGTCTGCTGCTGCGCGCGCGCGGCAACGAGGGCAACGAGCGGAGCAGCAGCACC 798
QY 658 CTGGACAGCGACGTGAACCCAGAGGACCCCGGCGCAGATGGCGTGTGGCGGTATCACC 717
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
799 GTCGGTACCGGGCAAGTACCCGAGCGTATCGCCGTCGCGCGCGTGGACAGCAGCAAC 858
QY 718 CTGGTGGGTGTGCCACCCCGTGAACGTGCGCGGAGCACTGCTCCTCCCGAGGGGAC 777
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
859 CAGCGCGGAGCTTACAGCAGCGTTCGGCCCGAGCTGGACGTATGGCCCGCGGCGTACG 918
QY 778 ACCCCAGTGTAGACAAGGGGAGGGGAGGTGGCCACCATCGCCAAACGGGAAGGTCAAC 837
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
919 ATCCAGAGCACCTTCCCGGGCAACAAGTACGGGCGCCTACAGCGGCACCAAGCATGGCCAGC 978
QY 838 CCG 840
Db |||||
979 CCG 981

RESULT 3
US-09-773-816-1
; Sequence 1, Application US/09773816
; Patent No. 6340774
; GENERAL INFORMATION:
; APPLICANT: Stanford University
; APPLICANT: Khosla, Chaitan
; TITLE OF INVENTION: NON-STEROIDAL ESTROGEN-RECEPTOR
; FILE REFERENCE: 28600-20210.00
; CURRENT APPLICATION NUMBER: US/09/773,816
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/243,458
; PRIOR FILING DATE: 2000-10-25
; PRIOR APPLICATION NUMBER: US 60/179,305
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSEQ for Windows Version 4.0
```

```
; SEQ ID NO 1
; LENGTH: 23673
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(23623)
; OTHER INFORMATION: n = A,T,C or G
US-09-773-816-1

Query Match      1.5%; Score 56.2; DB 4; Length 23673;
Best Local Similarity 50.2%; Pred. No. 0.01;
Matches 139; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 558 CACGGTGGTGACCAAGTTACCGTCTGCAACGCGCACGTCGTGTGTCATCTCCAGCATCCC 617
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
12896 CATTCGGGGATCGACCTGACCGCGGCACGCGCACTGGCTGAGCCGGCACGGCAACAC 12955
QY 618 CGCGGCCAGGACAGCGACTACCTCTCCCGGGAGATGTTCTCTGGACAGCGACGTGAACCC 677
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
12956 CACCGGATCGCGGACCGGTTCCGCGCGGAGCGGGTGTTCCTGGCCGGCAGCGGGCGCA 13015
QY 678 AGAGGACCCCGGCGCAGATGCGGTGCTGCGCGGTATCACCTGGTGGCTGTGCCACCG 737
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
13016 CGTGTCCCGCGCGGCGGCGGAGTTGCTGACCAACCGGGTGCACGACGGCAACCACT 13075
QY 738 CTGCAACGTGCGCGGAGCAACTGCTCCTCCGAGGGGACACCCAGTGTCTAGACAAGG 797
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
13076 GGGCTGGAAGCTCGGCGCGGGTGCCTGGGCTGGGCTCCGAGACCTGCTCGACAGCTA 13135
QY 798 GCAGGGGAGGTGGCCACCATCGCCAAACGGGAAGGTC 834
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
13136 CCAGGACGAGCGGTACCGGTCGCGGAGCGGATCATC 13172
```

```
RESULT 4
US-09-475-515-47
; Sequence 47, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 1836
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp140.mut8.modsFl62.delv1/v2
US-09-475-515-47

Query Match      1.5%; Score 55.2; DB 4; Length 1836;
Best Local Similarity 45.2%; Pred. No. 0.0076;
Matches 244; Conservative 0; Mismatches 293; Indels 3; Gaps 1;

QY 205 CCCAACGGGGCCAGGAGGACACCGGATGAAGAACGTGCGCGTCCCGGTGTACTGCCGC 264
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1090 CCCATCCGGGCCAGATCCGCTGACGAGCAACATCACCGGCTGTGTGACCGCGAC 1149
```

QY 265 CCTCTGGTGAGAGAGACCCACCATGAAGTGTGGTGTGCGCGGGCGGTCAACCTGAGC 324  
Db 1150 GCGGCAAGGAGATCAGCAACACCACCGAGATCTTCCGCCCGCGGGCGGACATGCGC 1209  
QY 325 GGGTGGAGGCCAATGAGGACACGCTGGGAATGGAGTCAAGCCAGCGCCAGCGCGAT 384  
Db 1210 GACAACTGGCGAGCGAGCTGTACAAGTACAAGGTGGTGAAGATCGAGCCCTGGGCGTG 1269  
QY 385 CCCCTGACCTGCGACCCGCGAAGGAGACGGCGAGCCCAAGAGCGCCACACGTCCTCCCGAG 444  
Db 1270 GCGCCACCATCGCCATCAGCAGCGTGGTGCGAGCGGAGAGAGCGCCGTGACCCCTGGG- 1328  
QY 445 AAGAAGAAGGCCAAGGAGCTCCCTGAAATGGAGCGCCACCTCCAGCCGGGTGTGGATCCTG 504  
Db 1329 --CGCCATGTTCTGGGCTTCTGGGCGCCCGCGGAGCACCATGGGCGCCCGCAGCCTG 1386  
QY 505 ACCAGCACCTGACCAACAGCAAGGTGGTGAATCGACGCGCAACAGCCAGCGCGGACCGGTG 564  
Db 1387 ACCCTGACCGTGCAGGCCCGCCAGCTGCTGAGCGGCATCGTGCGAGCAGCAGAACACCTG 1446  
QY 565 GTGGACCAGTTACCGTCTGCAACCGCGGAGATGTTCTCTGGACAGCGACGTGAACCCAGAGGAC 624  
Db 1447 CTGCGCGCCATCAGGCGCCAGCAGCACCTGCTGAGCTGACCGTGTGGGCAATCAAGCAG 1506  
QY 625 AGCGACAGCGACTACCTCCCGGGGAGATGTTCTCTGGACAGCGACGTGAACCCAGAGGAC 684  
Db 1507 CTGAGGCGCGGTGCTGGCGGTGAGCGCTACCTGAAGGACAGCAGCTGCTGGGCATC 1566  
QY 685 CCGGGCGCAGATGGCGTGTGCGCGGTATCACCTGGTGGGCTGTGCCACCCCGTGAAC 744  
Db 1567 TGGGGCTGCAGCGCAAGCTGATCTGCACCACCGCGGTGCCCTGGAACGCCAGCTGGAGC 1626

RESULT 5

US-09-475-515-46  
; Sequence 46, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 46  
; LENGTH: 1944  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: gp140.mut8.modSF162.delV2

US-09-475-515-46

Query Match 1.5%; Score 55.2; DB 4; Length 1944;  
Best Local Similarity 45.2%; Pred. No. 0.0077;  
Matches 244; Conservative 0; Mismatches 293; Indels 3; Gaps 1;  
QY 205 CCCAACGGGGCCAGGAGACACGCGGATGAAGACGTGCCGGTGGCGGTGTACTGCGGC 264  
Db 1198 CCCATCCGCGCCAGATCCGCTGCAGCAGCAACATCACGGGCTGTGCTGACCCGCGAC 1257  
QY 265 CCTCTGGTGAGAGGACCCACCATGAAGTGTGGTGTGCCGCGGGCGGTCAACCTGAGC 324

Db 1258 GCGGCAAGGAGATCAGCAACACCACCGAGATCTTCCGCCCGCGGGCGGACATGCGC 1317  
QY 325 GGGTGGAGGCCAATGAGGACGACGCTGGGAATGGAGTCAAGCCAGCGCCAGCGCGAT 384  
Db 1318 GACAACTGGCGAGCGAGCTGTACAAGTACAAGGTGGTGAAGATCGAGCCCTGGGCGTG 1377  
QY 385 CCCCTGACCTGCGACCCGCGAAGGAGACGGCGAGCCCAAGAGCGCCACACGTCCTCCCGAG 444  
Db 1378 GCGCCACCATCGCCATCAGCAGCGTGGTGCGAGCGGAGAGAGCGCCGTGACCCCTGGG- 1436  
QY 445 AAGAAGAAGGCCAAGGAGCTCCCTGAAATGGAGCGCCACCTCCAGCCGGGTGTGGATCCTG 504  
Db 1437 --CGCCATGTTCTGGGCTTCTGGGCGCCCGCGGAGCACCATGGGCGCCCGCAGCCTG 1494  
QY 505 ACCAGCACCTGACCAACAGCAAGGTGGTGAATCGACGCGCAACAGCCAGCGCGGACCGGTG 564  
Db 1495 ACCCTGACCGTGCAGGCCCGCGCAGCTGCTGAGCGGCATCGTGCGAGCAGCAGAACACCTG 1554  
QY 565 GTGGACCAGTTACCGTCTGCAACCGCGCAGCTGCTGTGATCTCCAGCATCCCGCGGCGC 624  
Db 1555 CTGCGCGCCATCGAGGCCCGCAGCACCTGCTGAGCTGACCGTGTGGGCATCAAGCAG 1614  
QY 625 AGCGACAGCGACTACCTCCCGGGGAGATGTTCTCTGGACAGCGACGTGAACCCAGAGGAC 684  
Db 1615 CTGAGGCGCGGTGCTGGCGGTGAGCGCTACCTGAAGGACAGCAGCTGCTGGGCATC 1674  
QY 685 CCGGGCGCAGATGGCGTGTGCGCGGTATCACCTGGTGGGCTGTGCCACCCCGTGAAC 744  
Db 1675 TGGGGCTGCAGCGCAAGCTGATCTGCACCACCGCGGTGCCCTGGAACGCCAGCTGGAGC 1734

RESULT 6

US-09-475-515-45  
; Sequence 45, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 45  
; LENGTH: 2025  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: gp140.mut8.modSF162

US-09-475-515-45

Query Match 1.5%; Score 55.2; DB 4; Length 2025;  
Best Local Similarity 45.2%; Pred. No. 0.0078;  
Matches 244; Conservative 0; Mismatches 293; Indels 3; Gaps 1;  
QY 205 CCCAACGGGGCCAGGAGACACGCGGATGAAGACGTGCCGGTGGCGGTGTACTGCGGC 264  
Db 1279 CCCATCCGCGCCAGATCCGCTGCAGCAGCAACATCACGGCCTGTGCTGACCCGCGAC 1338  
QY 265 CCTCTGGTGAGAGGACCCACCATGAAGTGTGGTGTGCCGCGGGCGGTCAACCTGAGC 324  
Db 1339 GCGGCAAGGAGATCAGCAACACCACCGAGATCTTCCGCCCGCGGGCGGACATGCGC 1398

|    |      |  |      |
|----|------|--|------|
| QY | 325  | GGTGGAGGCCCAATGAGGACGACGCTGGGAATGGAGTCAAGCCAGCGCCAGCCCGCAT     | 384  |
| Db | 1399 | GACAACTGGCGCAGGAGCTGTACAAGTACAAGGTGGTGAAGATCGAGCCCTGGGCGTG     | 1458 |
| QY | 385  | CCCTGACCTGGGACCGCGAAGGAGACGGGAGCCCAAGAGCGCCACACGTCCTCCCGAG     | 444  |
| Db | 1459 | GCCTCCACCATGCCATCAGCAGCGTGGTGAGAGCGAGAGCGCCGTGACCTTGGG-        | 1517 |
| QY | 445  | AAGAAGAAGGCCAAGGAGCTCCCTGAAATGGAGCCCACTCCAGCCGGGTGTGGATCCTG    | 504  |
| Db | 1518 | --CGCCATGTTCTGGGCTTCTTGGGCGCCGCGGCAGCACCATGGGCGCCGACGCTG       | 1575 |
| QY | 505  | ACCAGCACCTTGACCACCAAGCAAGGTGGTGATCATCGACGCCAACCAAGCCGGGCACGGTG | 564  |
| Db | 1576 | ACCTGACCGGTGAGGCCCGCCAGCTGCTGAGCGGCATCTGTCAGCAGCAGAACAAACCTG   | 1635 |
| QY | 565  | GTGACCAGTTCAACCGTCTGCAACGGCACGCTGTGTGCATCTCCAGCATCCCCCGCGCC    | 624  |
| Db | 1636 | CTGCGCGCCATCGAGGCCCAAGCAGCACCTGCTGCAGTGCCTGTGGGSCATCAAGCAG     | 1695 |
| QY | 625  | AGCGACAGCGACTACCCCTCCGGGGGAGATGTTCTTGACAGCGACGTGAACCCAGAGGAC   | 684  |
| Db | 1696 | CTGCAGGCCCGCGTGTGGCCGTGGAGCGCTACCTGAAGACCAGCAGCTGTGGGCATC      | 1755 |
| QY | 685  | CCGGGGCAGATGGCGTGTGGCCGGGTATCACCTTGGTGGGCTGTGCCACCCGCTGCAAC    | 744  |
| Db | 1756 | TGGGGCTGACGGGCAAGCTGATCTGCACACCGCCGTGCCCTGGAACGCCAGCTGGAGC     | 1815 |

## RESULT 7

US-09-000-016-3  
; Sequence 3, Application US/09000016  
; Patent No. 6143541  
; GENERAL INFORMATION:  
; APPLICANT: Akira ARISAWA et al.  
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC  
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV  
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
; STREET: 2033 K Street, N.W., #800  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20006  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/000,016  
; FILING DATE: January 30, 1998  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren M. Cheek, Jr.  
; REGISTRATION NUMBER: 33,367  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-721-8200  
; TELEFAX: 202-721-8250  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2539 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic DNA

```

; ORIGINAL SOURCE:
; ORGANISM: Streptomyces viridosporus
; STRAIN: A-914
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E
US-09-000-016-3

```

Query Match 1.5%; Score 54.8; DB 3; Length 2539;  
Best Local Similarity 44.9%;  
Matches 249; Conservative 0; Mismatches 302; Indels 3;  
Pred. No. 0.01;

|    |      |   |      |
|----|------|---|------|
| QY | 536  | TCATCGACGCGCAACACAGCCGGGCACGGTGGTGACCAAGTTACACCGTCTGCAACCGGCACG | 595  |
| Db | 1095 | TCACCGCGCGCCCGGCGCGCGACAAAGTGGGCCACGSCACCCACGCTCGCTCGATCG       | 1154 |
| QY | 596  | TGCTGTGCATCTCCAGCATCCCCGCGGCCACGACGACGCACTACCTCTCCCGGGAGATGT    | 655  |
| Db | 1155 | CGGCGGCACGGCGCCCACTCAAGGCAAGTACAAGGCGCTCGACCCCGGCCCGCA          | 1214 |
| QY | 656  | TCCTGGACACGACGTGAACCCAGAGGACCCGGGCGCAGATGGCGTGTCTGGCCGGTATCA    | 715  |
| Db | 1215 | TCCTCAACGGCAAGTCTCTCGACGACTCCGGTTTCGGCGACGACTCCGGCATCTCGCCG     | 1274 |
| QY | 716  | CCCTGTGGGTGTGCCACCCGCTGCAACGTGCCGCGGAGCAACTGCTCTCTCCCGAGGG      | 775  |
| Db | 1275 | GCATGGAGTGGCGCGCCGCGCAGGCGCGGACGTCGTCAACATGAGCCTGGGCGGCATGG     | 1334 |
| QY | 776  | ACACCCAGTGTAGACAAAGGGGCAGGGGAGGTGGCCACCATCGCCAAACGGGAAGTCA      | 835  |
| Db | 1335 | ACACACCGGAGACCGACCCGCTGGAGGCGCGGTGCACAAAGCTGTCCGCCAGAAAGGGCG    | 1394 |
| QY | 836  | ACCCGTCCCAGTCCACAGAGGAGGCCACAGAGGCCACGAGGTGCCAGACCTGGGCCCA      | 895  |
| Db | 1395 | TCCTGTTC---GCCATCGGGCCGGCAACGAGGGCCCGGAGTCGATCGGTTGCCCGGCA      | 1451 |
| QY | 896  | GCGAGCCAGAGACAGCCACATTGGGCCCCGGCCCTCTCAAGAGCACTGTTCACTGACC      | 955  |
| Db | 1452 | GCGCGACGCCGCCCTACCGTCCGGCGCCGTCGACGCAAGGACAAGCTCGCCGACTTCT      | 1511 |
| QY | 956  | CAGCCCCGACCCCGTCTCTGGCCCCCAGCCCTGGCAGGAGAACGGGCCAGAGCCCTGACA    | 1015 |
| Db | 1512 | CCTCCACGGCCCCCGCCCTCGGCGACGGCGCCATCAAGCCGGACGTCACCGCTCCCGCG     | 1571 |
| QY | 1016 | GCAGCAGCACACGGCCAGAGCCAGAGCCACGCGGGGACCCCAACGGGACAGGCAGT        | 1075 |
| Db | 1572 | TGGACATCACGGCGGCTCGGCGGAGGGCAACGACATCGGCCAGGAGTGGTGAGGAC        | 1631 |
| QY | 1076 | CTGCACCCACCATG  | 1089 |
| Db | 1632 | CGGCGGCTACATG   | 1645 |

## RESULT 8

US-09-514-340-3  
; Sequence 3, Application US/09514340  
; Patent No. 6361987  
; GENERAL INFORMATION:  
; APPLICANT: Akira ARISAWA et al.  
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC  
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-  
; ITS EXPRESSION PRODUCT  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
; STREET: 2033 K Street, N.W., #800  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20006  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb





```
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
; LENGTH: 1836
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp140.mut.modSF162.delV1/V2
US-09-475-515-41

Query Match      1.5%; Score 53.6; DB 4; Length 1836;
Best Local Similarity 45.0%; Pred. No. 0.017;
Matches 243; Conservative 0; Mismatches 294; Indels 3; Gaps 1;

QY 205 CCCAACGGGGCCAGGAGGACACCGCGATGAAGAACGTGCCGGTCCGGTGTAATCTGCGC 264
Db 1090 CCCATCCGCGGCAGATCCGCTGCAGCAGCAACATCACCGGCTGTGCTGACCCGCGAC 1149
QY 265 CCTCTGGTGAGAGGACCCCAACCATGAAGCTGTGGTGTGCCGGCGGCGTCAACTGAGC 324
Db 1150 GCGGCAAGGAGATCAGCAACACCAACCGAGATCTTCCGCCCGCGGCGGACATGCGC 1209
QY 325 GGTGGAGGCCCCAATGAGGACGCGCTGGGAATGAGTCAAGCCAGCGCCAGCGCGCAT 384
Db 1210 GACAACTGGCGCAGCGAGCTGTACAAGTACAAAGTGTGAAGTCAAGTGTGGCGGTG 1269
QY 385 CCCTGACCTGCGACCGCGAAGGAGACGGCGAGCCCAAGAGCGCCACACAGTCTCCCGAG 444
Db 1270 GCCCCACCAAGGCCAAGCGCGCGTGTGCAGCGGAGAGAGAGCGCCGTG---ACCTG 1326
QY 445 AAGAGAAAGGCCAAGGAGCTCCCTGAAATGGACGCCACCTCCAGCGGGTGTGGATCCTG 504
Db 1327 GCGGCCATGTTCTGGGCTTCTGGGCTTCTGGGCGCGCGCGAGCACCATGGGCGCCGCGCCTG 1386
QY 505 ACCAGCACCTGACCAACAGCAAGGTGGTGTATCATCGACGCCAACCCAGCGGGCACGGTG 564
Db 1387 ACCCTGACCGTGCAGGCCCGCCAGCTGTGAGCGGCATCGTGACGAGCAGAACACCTG 1446
QY 565 GTGGACCAAGTTACCGTCTGCAACGCGCACGTGTGTGCATCTCCAGCATCCCCCGCGCC 624
Db 1447 CTGCGGCCATCGAGGCCCGCAGCAGCACTGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 1506
QY 625 AGCAGACGCGACTACCTCCCGGGAGATGTTCTTGGACAGCGACGTGAACCCAGAGGAC 684
Db 1507 CTGAGGCCCGCGTGTGGCCGTGGAGCGCTACCTGAAGGACCGAGCAGTGTGGGCATC 1566
QY 685 CCGGGCGCAGATGGCGTGTGGCGGTATCACCCCTGTGGGCTGTGCCACCCCGTGC AAC 744
Db 1567 TGGGGTGTGACGGGCAAGCTGATCTGCACACCGCGCGTGTCCCTTGAACCGCAGCTGGAGC 1626
```

RESULT 12  
US-09-475-515-44  
; Sequence 44, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan

```
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 1836
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp140.mut7.modSF162.delV1/V2
US-09-475-515-44
```

```
Query Match      1.5%; Score 53.6; DB 4; Length 1836;
Best Local Similarity 45.0%; Pred. No. 0.017;
Matches 243; Conservative 0; Mismatches 294; Indels 3; Gaps 1;

QY 205 CCCAACGGGGCCAGGAGGACACCGCGATGAAGAACGTGCCGGTCCGGTGTAATCTGCGC 264
Db 1090 CCCATCCGCGGCAGATCCGCTGCAGCAGCAACATCACCGGCTGTGCTGACCCGCGAC 1149
QY 265 CCTCTGGTGAGAGGACCCCAACCATGAAGCTGTGGTGTGCCGGCGGCGTCAACTGAGC 324
Db 1150 GCGGCAAGGAGATCAGCAACACCAACCGAGATCTTCCGCCCGCGGCGGACATGCGC 1209
QY 325 GGTGGAGGCCCCAATGAGGACGCGCTGGGAATGAGTCAAGCCAGCGCCAGCGCGCAT 384
Db 1210 GACAACTGGCGCAGCGAGCTGTACAAGTACAAAGTGTGAAGTCAAGTGTGGCGGTG 1269
QY 385 CCCTGACCTGCGACCGCGAAGGAGACGGCGAGCCCAAGAGCGCCACACAGTCTCCCGAG 444
Db 1270 GCCCCACCAAGGCCAAGCGCGCGTGTGCAGCGGAGAGAGAGCGCCGTG---ACCTG 1326
QY 445 AAGAGAAAGGCCAAGGAGCTCCCTGAAATGGACGCCACCTCCAGCGGGTGTGGATCCTG 504
Db 1327 GCGGCCATGTTCTGGGCTTCTGGGCTTCTGGGCGCGCGCGAGCACCATGGGCGCCGCGCCTG 1386
QY 505 ACCAGCACCTGACCAACAGCAAGGTGGTGTATCATCGACGCCAACCCAGCGGGCACGGTG 564
Db 1387 ACCCTGACCGTGCAGGCCCGCCAGCTGTGAGCGGCATCGTGACGAGCAGAACACCTG 1446
QY 565 GTGGACCAAGTTACCGTCTGCAACGCGCACGTGTGTGCATCTCCAGCATCCCCCGCGCC 624
Db 1447 CTGCGGCCATCGAGGCCCGCAGCAGCACTGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 1506
QY 625 AGCAGACGCGACTACCTCCCGGGAGATGTTCTTGGACAGCGACGTGAACCCAGAGGAC 684
Db 1507 CTGAGGCCCGCGTGTGGCCGTGGAGCGCTACCTGAAGGACCGAGCAGTGTGGGCATC 1566
QY 685 CCGGGCGCAGATGGCGTGTGGCGGTATCACCCCTGTGGGCTGTGCCACCCCGTGC AAC 744
Db 1567 TGGGGTGTGACGGGCAAGCTGATCTGCACACCGCGCGTGTCCCTTGAACCGCAGCTGGAGC 1626
```

RESULT 13  
US-09-475-515-37  
; Sequence 37, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh

```
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 1944
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gpl40.modSF162.delV2
US-09-475-515-37
```

```
Query Match 1.5%; Score 53.6; DB 4; Length 1944;
Best Local Similarity 45.0%; Pred. No. 0.017;
Matches 243; Conservative 0; Mismatches 294; Indels 3; Gaps 1;

QY 205 CCCAACGGGGCCAGGAGGACACCGCGGATGAAGAACGTGCCGGTGCCTGCTACTGCGGC 264
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1198 CCCATCCGGCCAGATCCGCTGCAGCAGCACATCACCGGCTGCTGCTACCCCGGAC 1257
QY 265 CCTCTGGTGAGAGGACCCACCATGAAGTGTGGTGTGGCGGGGGGTCAACCTGAGC 324
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1258 GCGGCAAGGAGATCAGCAACACCCAGATCTTCCGCCCGGGCGGCGGACATGCGC 1317
QY 325 GGGTGGAGGCCCAATGAGGAGGACCGTGGGAATGGAGTCAAGCCAGCGCCGCGGAT 384
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1318 GACAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTGAAGATCGAGCCCTGGCGGTG 1377
QY 385 CCCCTGACTGCGACCGCGAAGGAGACGGCGAGCCCAAGAGCGCCCAACGTCTCCCGAG 444
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1378 GCCCCCAACCAAGGCCAAGCGCGGTGGTGCAGCGCGAGAA---GCGCGCGGTGACCCCTG 1434
QY 445 AAGAAGAGGCCCAAGGAGTCCCTGAAATGGACGCCACCTCCAGCCGGGTGTGGATCCTG 504
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1435 GCGCCATGTTCTCTGGGCTTCTGGGGCGCGCGGACCATGGGCGCCCGCAGCCTG 1494
QY 505 ACCAGCACCTGACCAACAGCAAGGTGGTGTATCATCGACGCCAACAGCGGGCACGGTG 564
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1495 ACCCTGACCGTGCAGGCCCGCCAGCTGCTGAGCGGCATCGTGCAGCAGCAACAACCTG 1554
QY 565 GTGACCAAGTTACCCGTCTGCAACGCGCAGCTGCTGTGCATCTCCAGATCCCCCGGGCC 624
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1555 CTGCGGCCATCGAGGCCCGGAGCACCCTGCTGACGCTGACCGTGTGGGGCATCAAGCAG 1614
QY 625 AGCGACAGGACTACCTCCCGGGGAGATGTTCTTGACAGCGACGTGAACCCAGAGGAC 684
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1615 CTGACGCCCGCGTGTGGCGGTGAGCGCTACCTGAAGGACACGACGCTGCTGGGCATC 1674
QY 685 CCGGCGCAGATGGCGTGTGGCGGTATCACCTGTGGGTGTGTCACCCCGCTGCAAC 744
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1675 TGGGCTGCAGCGGCAAGCTGATCTGCACACCGCGCTGCCCTGGAACGCCAGCTGGAGC 1734
```

```
RESULT 14
US-09-475-515-38
; Sequence 38, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEDEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
```

```
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 1944
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gpl40.modSF162.delV1/V2
US-09-475-515-38
```

```
Query Match 1.5%; Score 53.6; DB 4; Length 1944;
Best Local Similarity 45.0%; Pred. No. 0.017;
Matches 243; Conservative 0; Mismatches 294; Indels 3; Gaps 1;

QY 205 CCCAACGGGGCCAGGAGGACACCGCGGATGAAGAACGTGCCGGTGCCTGCTACTGCGGC 264
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1198 CCCATCCGGCCAGATCCGCTGCAGCAGCAACATCACCGGCTGCTGCTGACCCCGGAC 1257
QY 265 CCTCTGGTGAGAGGACCCACCATGAAGTGTGGTGTGGCGGGGGGTCAACCTGAGC 324
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1258 GCGGCAAGGAGATCAGCAACACCCAGATCTTCCGCCCGGGCGGCGGACATGCGC 1317
QY 325 GGGTGGAGGCCCAATGAGGAGGACCGTGGGAATGGAGTCAAGCCAGCGCCGCGGAT 384
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1318 GACAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTGAAGATCGAGCCCTGGCGGTG 1377
QY 385 CCCCTGACTGCGACCGCGAAGGAGACGGCGAGCCCAAGAGCGCCCAACGTCTCCCGAG 444
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1378 GCCCCCAACCAAGGCCAAGCGCGGTGGTGCAGCGCGAGAA---GCGCGCGGTGACCCCTG 1434
QY 445 AAGAAGAGGCCCAAGGAGTCCCTGAAATGGACGCCACCTCCAGCCGGGTGTGGATCCTG 504
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1435 GCGCCATGTTCTCTGGGCTTCTGGGGCGCGCGGACCATGGGCGCCCGCAGCCTG 1494
QY 505 ACCAGCACCTGACCAACAGCAAGGTGGTGTATCATCGACGCCAACAGCGGGCACGGTG 564
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1495 ACCCTGACCGTGCAGGCCCGCCAGCTGCTGAGCGGCATCGTGCAGCAGCAACAACCTG 1554
QY 565 GTGACCAAGTTACCCGTCTGCAACGCGCAGCTGCTGTGCATCTCCAGATCCCCCGGGCC 624
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1555 CTGCGGCCATCGAGGCCCGGAGCACCCTGCTGACGCTGACCGTGTGGGGCATCAAGCAG 1614
QY 625 AGCGACAGGACTACCTCCCGGGGAGATGTTCTTGACAGCGACGTGAACCCAGAGGAC 684
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1615 CTGACGCCCGCGTGTGGCGGTGAGCGCTACCTGAAGGACACGACGCTGCTGGGCATC 1674
QY 685 CCGGCGCAGATGGCGTGTGGCGGTATCACCTGTGGGTGTGTCACCCCGCTGCAAC 744
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1675 TGGGCTGCAGCGGCAAGCTGATCTGCACACCGCGCTGCCCTGGAACGCCAGCTGGAGC 1734
```

```
RESULT 15
US-09-475-515-40
; Sequence 40, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEDEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 11:10:58 ; Search time 21 Seconds  
(without alignments)  
1600.405 Million cell updates/sec

Title: US-10-019-495-9  
Perfect score: 3473  
Sequence: 1 MKNVPVPVCRPLVEKDPTM.....VLSKAERSHIIVQVSYTPE 651

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCITUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description        |
|------------|--------|-------------|--------|-------|--------------------|
| 1          | 1446.5 | 41.6        | 828    | 2     | US-08-993-228-21   |
| 2          | 146.5  | 4.2         | 985    | 5     | PCT-US96-03916-6   |
| 3          | 146.5  | 4.2         | 985    | 5     | PCT-US96-03916-66  |
| 4          | 143    | 4.1         | 490    | 3     | US-09-109-841-2    |
| 5          | 134    | 3.9         | 406    | 3     | US-08-582-740-68   |
| 6          | 134    | 3.9         | 406    | 3     | US-09-109-879-68   |
| 7          | 133    | 3.8         | 382    | 3     | US-08-582-740-70   |
| 8          | 133    | 3.8         | 382    | 3     | US-09-109-879-70   |
| 9          | 121.5  | 3.5         | 2441   | 1     | US-08-194-468-2    |
| 10         | 121.5  | 3.5         | 2441   | 3     | US-08-961-739-2    |
| 11         | 121.5  | 3.5         | 2441   | 4     | US-09-514-247A-8   |
| 12         | 121.5  | 3.5         | 2441   | 4     | US-09-686-316-2    |
| 13         | 119.5  | 3.4         | 1231   | 4     | US-08-714-741-41   |
| 14         | 119    | 3.4         | 3892   | 4     | US-09-328-352-5503 |
| 15         | 118.5  | 3.4         | 1038   | 4     | US-07-757-022B-74  |
| 16         | 118.5  | 3.4         | 1270   | 4     | US-07-757-022B-44  |
| 17         | 118.5  | 3.4         | 1311   | 4     | US-07-757-022B-42  |
| 18         | 118    | 3.4         | 941    | 4     | US-07-757-022B-14  |
| 19         | 118    | 3.4         | 1022   | 4     | US-07-757-022B-84  |
| 20         | 118    | 3.4         | 1049   | 4     | US-07-757-022B-58  |
| 21         | 118    | 3.4         | 1140   | 4     | US-07-757-022B-104 |
| 22         | 118    | 3.4         | 1313   | 4     | US-07-757-022B-142 |
| 23         | 118    | 3.4         | 1314   | 4     | US-07-757-022B-50  |
| 24         | 118    | 3.4         | 1320   | 4     | US-07-757-022B-46  |
| 25         | 118    | 3.4         | 1320   | 4     | US-07-757-022B-60  |
| 26         | 118    | 3.4         | 1320   | 4     | US-10-164-595-58   |
| 27         | 118    | 3.4         | 1354   | 4     | US-07-757-022B-48  |

|    |       |     |      |   |                   |                   |
|----|-------|-----|------|---|-------------------|-------------------|
| 28 | 118   | 3.4 | 1361 | 4 | US-07-757-022B-40 | Sequence 40, Appl |
| 29 | 118   | 3.4 | 1363 | 4 | US-07-757-022B-52 | Sequence 52, Appl |
| 30 | 118   | 3.4 | 1404 | 4 | US-07-757-022B-2  | Sequence 2, Appli |
| 31 | 118   | 3.4 | 1404 | 4 | US-07-757-022B-62 | Sequence 62, Appl |
| 32 | 118   | 3.4 | 1404 | 4 | US-10-164-595-78  | Sequence 78, Appl |
| 33 | 117   | 3.4 | 1426 | 3 | US-09-136-574A-43 | Sequence 43, Appl |
| 34 | 117   | 3.4 | 8991 | 4 | US-08-714-741-32  | Sequence 32, Appl |
| 35 | 116.5 | 3.4 | 1964 | 2 | US-08-790-912-3   | Sequence 3, Appli |
| 36 | 116.5 | 3.4 | 2052 | 2 | US-08-790-912-2   | Sequence 2, Appli |
| 37 | 116   | 3.3 | 1384 | 3 | US-08-976-255-11  | Sequence 11, Appl |
| 38 | 115.5 | 3.3 | 1719 | 2 | US-08-459-568-4   | Sequence 4, Appli |
| 39 | 115.5 | 3.3 | 1719 | 2 | US-08-399-411-4   | Sequence 4, Appli |
| 40 | 115.5 | 3.3 | 1719 | 3 | US-08-516-859A-4  | Sequence 4, Appli |
| 41 | 115.5 | 3.3 | 1719 | 4 | US-09-586-472-4   | Sequence 4, Appli |
| 42 | 114.5 | 3.3 | 1719 | 4 | US-09-528-706-4   | Sequence 4, Appli |
| 43 | 114.5 | 3.3 | 325  | 4 | US-09-510-031A-2  | Sequence 2, Appli |
| 44 | 114.5 | 3.3 | 641  | 3 | US-08-961-083-160 | Sequence 160, App |
| 45 | 114.5 | 3.3 | 641  | 4 | US-09-536-784-160 | Sequence 160, App |

ALIGNMENTS

RESULT 1  
US-08-993-228-21  
; Sequence 21, Application US/08993228  
; Patent No. 5976838  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: Lavallie, Edward R.  
; APPLICANT: Racie, Lisa A.  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Agostino, Michael J.  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
; TITLE OF INVENTION: ENCODING THEM  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/993,228  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sprunger, Suzanne A.  
; REGISTRATION NUMBER: 41,323  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 498-8284  
; TELEFAX: (617) 876-5851  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 828 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-993-228-21

Query Match 41.6%; Score 1446.5; DB 2; Length 828;  
Best Local Similarity 57.7%; Pred. No. 1.8e-126;  
Matches 300; Conservative 65; Mismatches 124; Indels 31; Gaps 10;

QY 1 MKNVPVYCRPLVEKDPMTKMLWCAAGVNLGWRPNEDDAGNGVKPAP-GRDPLTCDREG 59  
Db 314 MKNLPVPVYLRPLDEKDTSMKMLWCAVGNLSGGKTR--DGGSVVGASVFYKDVAGLDTEG 371  
QY 60 DGEPKSAHTSPE-----KKKAKELPEMDATSSRWILTSTLTTSKVVIIDANQPGTVVD 113  
Db 372 SKORSASQSSLDKLDQELKEQKELKNQELSSLVWICTSHSATKVLIIIDAVQPGNILD 431  
QY 114 QFTVCNAHVLCISSIPAAASDSDYPPGEMFLDSD-----VNPEDGADGVLAGIT 162  
Db 432 SFTVCNSHVLCIASVPGARETDYPAGEDLSESGQVDKASLCGSMTSNSSABTDSLLGGIT 491  
QY 163 LVGC-ATRCNVPRSNCSRGDTPVLDKGQGEVATIANGKVNPSQST-EEATEATEVPDPG 220  
Db 492 VVGCSAEGVTGAATSPSTNGASPVMDKPPMEA--ENSEVDENVPTAEEATEATE-GNAG 548  
QY 221 PSEPETATLRPGPLTEHVFTDPAPTP-----SSGPQPGSENGPEPDSSSTRP-EPEPSGD 274  
Db 549 SAEDTVDISQTVGVYTEHVFTDPLGVQIPEDLSPVYQSSNDSDAYKDQISVLENEQDLVRE 608  
QY 275 PTGAGSSAAPTMLWGAQNGWLYVHSAVANWKCLHSIKLSDVLSLVHVKGRLVALADG 334  
Db 609 EAQKMSSLLPTMLWGAQNGCLYVHSSVAQWRKCLHSIKLSDLSIVHVKGIVLVALADG 668  
QY 335 TLAIHRGEDGWDLSNYHLMDLGHPHHSIRCMVAVYDRVWCGYKNKVHVIOPKTMQIEK 394  
Db 669 TLAIHRGVDGWDLSNYHLLDLGRPHHSIRCMVTVVHDKVWCGYRNKIYVVOVKAMKIEK 728  
QY 395 SFDAHPRRESQVRQLAWIGDGVWVSIIRLDSTLRLYHAHTHQHLDQVDIEPYVSKMLGTGK 454  
Db 729 SFDAHPRKESQVRQLAWVGDVWVSIIRLDSTLRLYHAHTYQHLQDQVDIEPYVSKMLGTGK 788  
QY 455 LGFSFVRITALLVAGSRLVWGTGNGVWISIPLETETVWLHR 494  
Db 789 LGFSFVRITALLVMSNRLVWGTGNGVWIIISIPLETESKYIFR 828

RESULT 2  
PCT-US96-03916-6  
; Sequence 6, Application PC/TUS9603916  
; GENERAL INFORMATION:  
; APPLICANT: Wild, Martha A.  
; APPLICANT: Cochran, Mark D.  
; TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/03916  
; FILING DATE: 23-MAR-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/126,597  
; FILING DATE: 24-SEP-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 39116-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0525

; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 985 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US96-03916-6  
  
Query Match 4.2%; Score 146.5; DB 5; Length 985;  
Best Local Similarity 19.7%; Pred. No. 0.00018;  
Matches 118; Conservative 83; Mismatches 234; Indels 165; Gaps 28;  
  
QY 3 NVPVPVYCRPLVEKDPMTKMLWCAAGVNLGWRPNEDDAGNGVKP--APGRDPLTCDREGD 60  
Db 394 SVPVQETDRTLSTTTLPL-----TPGESE--NTLFTTAPGISTET----- 434  
QY 61 GEPKSAHTSPEKKAK-----ELPEMDATSSR-----VWILTSTLTTSKVVIIDANQPGT 110  
Db 435 --PSAAHETTQTQSAETVVFTQSPSTESATARSQSQEPWYFTQTPSTEQAAALTQTQIAET 492  
QY 111 VVDQFTVCNAHVLCISSIPAA-----SDSDYPPGEMFLDSDVNPD-----PGADGVL- 158  
Db 493 EALFTQTPSAEQMTFTQTPGAETEAPAQTPSTIP--EIFQSRSTPPETARAPSAPEVF 550  
QY 159 ----AGITLVGCATRCNVPRSNCSRGDTPVLDKGQ--GEVATIANGKVNPSQSTEEATE 212  
Db 551 TQSSSTVTEVFTQTPSTVPKTLSSSTEPALFTRTQSAGTEAFTQTSSAEPDTMRTQSTE 610  
QY 213 A---TEVPDGPSEPETATLRPGPLTEHVFTDPAP-----TPSSGPQPGSEN 256  
Db 611 THFFTQAPSTVPKATQTPSTPEVLTQSPSTPEVPFTRTLGAPEITQTPSAAPEVYTR- 669  
QY 257 GPEPDSSSTRPE-----PEPSGDPTGAGS-----SAAPTMWLGAQNGWLYVH 298  
Db 670 -----SSSTMPETAQSTPLASQNPTSSGTGTHNTEPRTPYVQTPHTQKLYTENKTLSP 724  
QY 299 SAVANWKKCLHSIKLSDVLSLVHVK-----GRV-----L 328  
Db 725 TVVSEF---HEMSTAESQTPLLDVKIVEKFSNDGEVATATCVTSKSPYRVETNWKVDL 780  
QY 329 VALAD-----GTLAIFHRGEDGQWDLNSYHLMDLGHPHHSIRCMVAV--YDRVWCGYKNK 381  
Db 781 VDMDEISGNSPAGVFNSE--KWQKQLYRYRVTGRTSVQMLMCLSCSHSPEPYCLFDTS 838  
QY 382 VHVQPCKTMQIEKSFDAHPRR-----ESQVRQLAWIGDGVWVSIIRLDSTLRLYHAHT 433  
Db 839 L-IAREKDIAPELYFTSDPQTAYCTITLPSGVVPRFEWSLNNVSLPEYLTATTVVSHTAG 897  
QY 434 HQHLQDQVDI---EPYVSKMLGTGKLGFSFVRITALLVAGSRL-----WVGTGNG 479  
Db 898 QSTVWKKSSARAGEAWIS-----GR-GGNIYECTVLISDGRVTRTKRCLNTWTIAVENG 951  
  
RESULT 3  
PCT-US96-03916-66  
; Sequence 66, Application PC/TUS9603916  
; GENERAL INFORMATION:  
; APPLICANT: Wild, Martha A.  
; APPLICANT: Cochran, Mark D.  
; TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0525

SOFTWARE: PatentIn Release #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/03916  
FILING DATE: 23-MAR-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/126,597  
FILING DATE: 24-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 39116-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 985 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US96-03916-66

Query Match 4.2%; Score 146.5; DB 5; Length 985;  
Best Local Similarity 19.7%; Pred. No. 0.00018;  
Matches 118; Conservative 83; Mismatches 234; Indels 165; Gaps 28;  
QY 3 NVPVPVYCRPLVEKDPMTKMLWCAAGVNLGWRPNEDDAGNGVKP--APGRDPLTCDREGD 60  
Db 394 SVPVQETDRLSTPLTLPL-----TPGESE--NTLPFTTAGISTET----- 434  
QY 61 GEPKSAHTSPEKKKAK-----ELPEMDATSSR-----VWILTSTLTTSKVVIIDANQPGT 110  
Db 435 --PSAAHETTQTSAAETVVFTQSPSTESFARSSQEPWYFTQTPSTEQAALTQTQIAET 492  
QY 111 VVDQFTVCNAHVLCISSIPAA-----SDSDYPPGEMFLSDVNPED-----PGADGVL- 158  
Db 493 EALFTQPSAEQMTFTQTPGAETEAPAQTPSTIP--EIFTQSRSTPPTETARAPSAPEVF 550  
QY 159 ----AGITLVGCATRCNVPNSCNRGDTPLVDKGQ--GEVATIANGKVNPSQSTEEATE 212  
Db 551 TQSSSTVEVFTQTPSTVPKNTLSSSTEPALFTRTQSAAGTEAFTQTSSAEPDTRMTQSTE 610  
QY 213 A---TEVPDPGPSEPETATLRPGPLTEHVFTDPAP-----TPSSGPGQPSGN 256  
Db 611 THFTQAPSTVPKATQTPSTEPEVLTSQSPSTEPVFPFTRTLGAPEITQTPSAAPEVYTR- 669  
QY 257 GPEPDSSSTRPE-----PEPSGDPTGAGS-----SAAPTMLGAQNGWLYVH 298  
Db 670 ----SSSTMPETAQSTPLASQNPSTSGTGTHNTEPRTPVQVTPHTQKLYTENKTLSP 724  
QY 299 SAVANWKCLHSIKLSDVLSLVHVK-----GRV-----L 328  
Db 725 TVVSEF-----HEMSTAESQTPLLDVKIVEVKFSNDGEVTATCVSTVKSPYRVEITNWKVDL 780  
QY 329 VALAD-----GTLAIFHRGEDGQWDLNSYHLMDLGHPHHSIRCMVAV--YDRVWCGYKNK 381  
Db 781 VDVMDEISGNSPAGVFNSE--KWQKQLYRYRVTGRTSVQLMCLSCSTSHSPEPYCLFDT 838  
QY 382 VHVIQPKTMQIEKSFDAHPRR-----ESQVRQLAWIGDGVWVSIKLDSTLRLYHAHT 433  
Db 839 L-IAREKXDIAPELYFTDPQTAYCTITLPSGVVPRFESLNNVSLPEYLTAITVVSHTAG 897  
QY 434 HQHLQDVDI---EPYVSKMLGTGKLGFSFVRITALLVAGSRL-----WVGTGNG 479  
Db 898 QSTVWKKSSARAGEAWIS-----GR-GGNIYECTVLISDGTTRVTRTKERCLTNTWIAVENG 951

RESULT 4  
US-09-109-841-2  
; Sequence 2, Application US/09109841  
; Patent No. 6207436  
; GENERAL INFORMATION:

APPLICANT: Bjornvad, Mads E.  
APPLICANT: Hatakeyama, Mariko  
APPLICANT: Schulein, Martin  
APPLICANT: Nielsen, Jack B.  
TITLE OF INVENTION: Endo-Beta-1, 4-Glucanases From  
TITLE OF INVENTION: Saccharothrix  
FILE REFERENCE: 5195.200-US  
CURRENT APPLICATION NUMBER: US/09/109,841  
CURRENT FILING DATE: 1998-07-02  
EARLIER APPLICATION NUMBER: 0812/97  
EARLIER FILING DATE: 1997-07-04  
EARLIER APPLICATION NUMBER: 0846/97  
EARLIER FILING DATE: 1997-07-11  
EARLIER APPLICATION NUMBER: 60/053,506  
EARLIER FILING DATE: 1997-07-23  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 490  
TYPE: PRT  
ORGANISM: Saccharothrix australiensis  
US-09-109-841-2

Query Match 4.1%; Score 143; DB 3; Length 490;  
Best Local Similarity 19.7%; Pred. No. 0.00013;  
Matches 94; Conservative 40; Mismatches 153; Indels 190; Gaps 18;  
QY 195 TIANGKVNPSQSTEEATEATEVDPDPGSEETATLRPGPLTEHVFTDPAPTSSGPGQGS 254  
Db 140 TSCNGAVG-GPTTEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPT 191  
QY 255 ENGPEPDSSSTRPEPEPSGDPTGAGSSAAPT-----MWLGA-----QNG 293  
Db 192 EPTPEP-----TPEPTPEPTMPVQAGQHFVDTTNSYRAWQAASGDKDLLAKIALTPQAY 248  
QY 294 WLYVHSAVANWKKCLHSIKLSDVLSLVHVKGR--VLVALADGTALIFHRGEDGQWDLN 351  
Db 249 W-----VGNWNEASHAQOEVRDITSAAGAAAGRTAVLVVYA-----IPGRD----- 288  
QY 352 YHLMDLGHPHHSIRCMVAVVYDRVWCGYKNKVHVIOPKTMQIEKSFDAHPRRESQVRQLAW 411  
Db 289 -----CGQHSSTGGV--STSEYAAQWIDT----- 308  
QY 412 IGDGV----WVSIRLDSTRLYHAHTQHQLQDVVDIEPYVSKMLGTG-KLGFSFVRITALL 466  
Db 309 VAQGI VGNPFWVL-----DPDALPMLGDCDQGGDRVGFVKYAAKSLT 350  
QY 467 VAGSRLLWVGTVGNGVVISIPLTETVVLHRLGQLLGLRANKTSPTSGEGARPGGIIHVYDGS 526  
Db 351 AKGARVYIDAGHSAWLS----- 367  
QY 527 SDRASSFIPYCSMAQAQLCFHGRDAVKFFVSV-----GNVLATLNGS---V 572  
Db 368 -----PSEANRLNQIGFEDAVGFSINVSNYRTTAESKWTGQVSVQLTGKPKFV 416  
QY 573 LDSPREGPGPAA-----PASEVEGQKLRNVLSGGEGYIDFRIGDGEDDETEEGAG 624  
Db 417 IDTSRNGNGPSGSEWNCNPSGRALGERPTLVNDRSGLDALLWIKLPGESDGCACNGGPG 473

RESULT 5  
US-08-582-740-68  
; Sequence 68, Application US/08582740  
; Patent No. 6037324  
; GENERAL INFORMATION:  
APPLICANT: Schwender, Charles F.  
APPLICANT: Shroff, Hitesh N.  
TITLE OF INVENTION: Inhibitors of MadCAM-1-Mediated  
TITLE OF INVENTION: Interactions and Methods of Use Therefor  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive



; APPLICANT: Schwender, Charles F.  
; APPLICANT: Shroff, Hitesh N.  
; TITLE OF INVENTION: Inhibitors of MadCAM-1-Mediated  
; TITLE OF INVENTION: Interactions and Methods of Use Therefor  
; NUMBER OF SEQUENCES: 70  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/582,740  
; FILING DATE: 04-JAN-1996  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brook, David E.  
; REGISTRATION NUMBER: 22,592  
; REFERENCE/DOCKET NUMBER: LKS95-12  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 861-6240  
; TELEFAX: (617) 861-9540  
; INFORMATION FOR SEQ ID NO: 70:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 382 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-582-740-70

Query Match 3.8%; Score 133; DB 3; Length 382;  
Best Local Similarity 22.1%; Pred. No. 0.00073;  
Matches 91; Conservative 38; Mismatches 138; Indels 144; Gaps 24;  
QY 5 PVPVYCRPL-VEKDPMTKLWCA-AGVNLGWRPNEDDAGNGVKPAPGRDPLTCDREGDGE 62  
Db 31 PEPVVAVALGASRQLTCLACADRGASVQ-WRGLDTSLG-AVQSDTGRSVLTVRN----- 83  
QY 63 PKSAHTSPEKKAKELPEMDATSSR-----VWILTSTLTTSKVVIID 104  
Db 84 -----ASLSAAGTRVCVSCGGRTFQHTVQLLVYAFDPQLTVSPAALV- 126  
QY 105 ANQPGTVVDQFTVCNAH-----VLCISSIPAASDSYPPGEMFLDSDVNPED---PG 153  
Db 127 ---PG---DPEVACTAHKVTVPDPNALSFSLLVGGQELE---GAQALGPEVQEEEEEPQG 177  
QY 154 ADGVLAGIT-----LVGCATRCNVPRNSCSRGTTPVLDKQGGEVATIANG 199  
Db 178 DEDVLFVRVTERWRRLPPLGTPVPPALYCOATMRLPGLSLHROAIPVLH----- 225  
QY 200 KVNPSQSTEEATEATEVPD-PGPSEPETATLRPGPLTEHVFTDPAP-----TPSSGPQPGS 254  
Db 226 --SPTSPEPPDTSPESPDTSPEPDTSQEP-PDT-----TSPEPPDKTSPEPAPOQGS 278  
QY 255 ENGP-EPDSSST-RPEPEPSGD-----PTG-----AGSSAAPTMLGAQNGWLYVHSAVA 302  
Db 279 THTPRSPGSTRTRRRPEISQAGPTQGEVPTGSSKPAAGDQLPALW-----TSSAVL 329  
QY 303 N-----WKKCLH-----SIKLSDVLSLVHVKGRVLVALA 332  
Db 330 GLLLLALPTYHLWKRCHLAEDDTHPPASRLLPQVSAWAGLRGTGQVGIS 380

RESULT 8  
US-09-109-879-70  
; Sequence 70, Application US/09109879  
; Patent No. 6274556

; GENERAL INFORMATION:  
; APPLICANT: Schwender, Charles F.  
; APPLICANT: Shroff, Hitesh N.  
; TITLE OF INVENTION: INHIBITORS OF MADCAM-1-MEDIATED  
; TITLE OF INVENTION: INTERACTIONS AND METHODS OF USE THEREFOR  
; NUMBER OF SEQUENCES: 89  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02421  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/109,879  
; FILING DATE: 02-JUL-1998  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US97/00291  
; FILING DATE: 03-JAN-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/582,740  
; FILING DATE: 04-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brook, David E.  
; REGISTRATION NUMBER: 22,592  
; REFERENCE/DOCKET NUMBER: LKS95-12A2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781) 861-6240  
; TELEFAX: (781) 861-9540  
; INFORMATION FOR SEQ ID NO: 70:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 382 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-109-879-70

Query Match 3.8%; Score 133; DB 3; Length 382;  
Best Local Similarity 22.1%; Pred. No. 0.00073;  
Matches 91; Conservative 38; Mismatches 138; Indels 144; Gaps 24;  
QY 5 PVPVYCRPL-VEKDPMTKLWCA-AGVNLGWRPNEDDAGNGVKPAPGRDPLTCDREGDGE 62  
Db 31 PEPVVAVALGASRQLTCLACADRGASVQ-WRGLDTSLG-AVQSDTGRSVLTVRN----- 83  
QY 63 PKSAHTSPEKKAKELPEMDATSSR-----VWILTSTLTTSKVVIID 104  
Db 84 -----ASLSAAGTRVCVSCGGRTFQHTVQLLVYAFDPQLTVSPAALV- 126  
QY 105 ANQPGTVVDQFTVCNAH-----VLCISSIPAASDSYPPGEMFLDSDVNPED---PG 153  
Db 127 ---PG---DPEVACTAHKVTVPDPNALSFSLLVGGQELE---GAQALGPEVQEEEEEPQG 177  
QY 154 ADGVLAGIT-----LVGCATRCNVPRNSCSRGTTPVLDKQGGEVATIANG 199  
Db 178 DEDVLFVRVTERWRRLPPLGTPVPPALYCOATMRLPGLSLHROAIPVLH----- 225  
QY 200 KVNPSQSTEEATEATEVPD-PGPSEPETATLRPGPLTEHVFTDPAP-----TPSSGPQPGS 254  
Db 226 --SPTSPEPPDTSPESPDTSPEPDTSQEP-PDT-----TSPEPPDKTSPEPAPOQGS 278  
QY 255 ENGP-EPDSSST-RPEPEPSGD-----PTG-----AGSSAAPTMLGAQNGWLYVHSAVA 302  
Db 279 THTPRSPGSTRTRRRPEISQAGPTQGEVPTGSSKPAAGDQLPALW-----TSSAVL 329  
QY 303 N-----WKKCLH-----SIKLSDVLSLVHVKGRVLVALA 332



QY 229 LRPGLTEHVFTDPAFTSSGQPGSENGPE-PDSSSTRPEPEPSGDPGTGAGSSAAPTMM 287  
Db 865 AAGMPSLQHP-TAPGMTP---PQAAAPTQSTPVSSGQTPTPTPGSVFSAAQSTPTTVQ 920  
QY 288 LGAQ 291  
Db 921 AAAQ 924

RESULT 12  
US-09-686-316-2  
; Sequence 2, Application US/09686316  
; Patent No. 6646115  
; GENERAL INFORMATION:  
; APPLICANT: Montminy, Marc R.  
; TITLE OF INVENTION: Methods for Treating Diabetes Mellitus  
; FILE REFERENCE: SALK1650-1  
; CURRENT APPLICATION NUMBER: US/09/686,316  
; PRIOR APPLICATION DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: US/08/961,739  
; PRIOR FILING DATE: 1997-10-31  
; PRIOR APPLICATION NUMBER: US 194,468  
; PRIOR FILING DATE: 1994-02-10  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 2441  
; TYPE: PRT  
; ORGANISM: Mus  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(2441)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-09-686-316-2

Query Match 3.5%; Score 121.5; DB 4; Length 2441;  
Best Local Similarity 27.2%; Pred. No. 0.17;  
Matches 50; Conservative 24; Mismatches 79; Indels 31; Gaps 9;  
QY 125 ISSIPAAASDSDYPPGEMFLDSVNPEDPGADGVLAGITLVGCATRCN-VPRSNCSRGD 182  
Db 755 MASVPGMAIS---PSRM-----PQPPNMGMTHANNIMAQAPTQNFQFPSSSGA 804  
QY 183 TPVLDKGGQEVATION-----GKVNPSQSTEEATEATEVPDP-----GPSEPETAT 228  
Db 805 MSVNSVGMGQPAAGVSGQEPGAALPNLMLAPQASQLPCPPVTQSPHLHTPPAST 864  
QY 229 LRPGLTEHVFTDPAFTSSGQPGSENGPE-PDSSSTRPEPEPSGDPGTGAGSSAAPTMM 287  
Db 865 AAGMPSLQHP-TAPGMTP---PQAAAPTQSTPVSSGQTPTPTPGSVFSAAQSTPTTVQ 920  
QY 288 LGAQ 291  
Db 921 AAAQ 924

RESULT 13  
US-08-714-741-41  
; Sequence 41, Application US/08714741  
; Patent No. 6500613  
; GENERAL INFORMATION:  
; APPLICANT: Briles, David E.  
; APPLICANT: McDaniel, Larry S.  
; APPLICANT: Swiatlo, Edwin  
; APPLICANT: Yother, Janet  
; APPLICANT: Crain, Marilyn J.  
; APPLICANT: Hollingshead, Susan  
; APPLICANT: Tart, Rebecca  
; APPLICANT: Brooks-Walter, Alexis  
; TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,  
; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,  
; TITLE OF INVENTION: PORTIONS AND PRODUCTS  
; NUMBER OF SEQUENCES: 47

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis, Morris & Safford, P.C.  
; STREET: 530 Fifth Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/714,741  
; FILING DATE: 16-SEP-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Frommer Esq., William S.  
; REGISTRATION NUMBER: 25,506  
; REFERENCE/DOCKET NUMBER: 454312-2460  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 840-3333  
; TELEFAX: (212) 840-0712  
; INFORMATION FOR SEQ ID NO: 41:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1231 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: amino acid  
US-08-714-741-41

Query Match 3.4%; Score 119.5; DB 4; Length 1231;  
Best Local Similarity 25.6%; Pred. No. 0.088;  
Matches 33; Conservative 14; Mismatches 53; Indels 29; Gaps 5;  
QY 203 PSQSTEEATEATEVPDPGPSEPETATLRPGPLT---EHVFTDPAFTSSGQPGSENGPE 259  
Db 652 PEKPAEKPAPEKPAPEKPAPEKPAPEKPAPEKPAPEKPAPEKPAPEKPAPEKPAPEKPA 707  
QY 260 PDSSSTRPEPEPSGD-----PTGAGSSAAPTMM-----WLGAGNGWLYVH-- 298  
Db 708 PEKPAEQPKPEKTDQQAEDYARRPEKPAPEKPAPEKPAPEKPAPEKPAPEKPAPEKPA 767  
QY 299 -SAVANWKK 306  
Db 768 GSMATGWSE 776

RESULT 14  
US-09-328-352-5503  
; Sequence 5503, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 5503  
; LENGTH: 3892  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-5503  
Query Match 3.4%; Score 119; DB 4; Length 3892;  
Best Local Similarity 20.4%; Pred. No. 0.61;  
Matches 150; Conservative 76; Mismatches 286; Indels 224; Gaps 34;  
QY 8 VYCRPLVEKOPTMKLWCAAGVNLSCWRPNEDDAGNGVKPAPGRDPLTCDREGDGEPKSAH 67

```
Db 144 VIYQPISSIEPL--LYHDAGVNPWLWAAIPLVAGGIIAAASNHD-----SNDDSSAPAD 195
QY 68 TSPEKKKAKELPEMDATSSRWILSTLTTSKWWIIDANQGTWVDQFTVCNAHVLCISS 127
Db 196 TTP-----PSTDGVTFSV-----DPVTSNDVINASEAGNV-----TITGV-----LKN 234
QY 128 IPAASDSYPPGEMFLDSVNPEDPGADGVLAGITLVGCAPFC-----NVPRSNCSRG 181
Db 235 IPA-----DAANTAVTVVINGWTYNATVDKAAAGTWTVSPGSLVADA 277
QY 182 DTPVLDKQGGEVATIANGKVNPSQSTEEATEPVP--DP-GPSEPETATLRPGPLTEHV 238
Db 278 DKTIDAKVTFTDAAGNSSSVNDTQTYTLDTTAPNAPVIDPVNGTDPITGTAEPSGTVTVT 337
QY 239 FTD-PAPTPSSG-----POPGSENGPEPDSSTRPEPEPSGDPPTGAGSSAAPTMWLGA 290
Db 338 YPDGSTKTVVAGPDGTWTVPNPGLNDGDEVTAATDPAGNTSGPATAVVDVAPTV---- 393
QY 291 QNGWLYVHSAVANWKKCLHSIKLKDSVLSL---VHVKGRLVALADGTLAIFHRGEDGQW 347
Db 394 -----ALDDVLTNDSPTALGTGVNDPTATVVVNVVDGVDYPVAVNGDGTW 437
QY 348 DLSNYHMLDLGHPHHSIRCM-----VVYDRVMCGYKKNKHVIOQKTMQIEKS 395
Db 438 TLADNTLPTLADGPHTITVTATDAAGNVGTDGTW-----TVDTAAPNTAGVTFT 487
QY 396 FDAHPRESQVQLAWIGD---GWVVSIRLSDTLRLYHAHQHLQDQVDIEPVVSKMLG 451
Db 488 IDS-VTADNVINASEAAGNVTTIGVLKNIPADAT---NTAVTVVINGVTYNATVVDKTAG 542
QY 452 TGKLGFSFVRITALLVAGSRLWVGTVNGVVISIPLTETV---VLHRGQLLGL----- 500
Db 543 TWTVS-----VPGSLVADADKTIIDAKVTFTDAAGNSSTVNDTQIYTLDTAAPAA 592
QY 501 ----RANKTSPTSCEGARPGGIIHVV-----GDD-----S 526
Db 593 PVIDPVNGTDPITGT-AEPGSGTIVTYTPNGDGTATVVAGPDGWSVNPGLNDGDEVEAIA 651
QY 527 SDBAASSFIPYCSMAQAQLCFHGHRAVKFFV-----SVPGNVLATLNGSVLD 574
Db 652 TDPAGNPSLPGTATVDA--VGPNTDGVNFTVDSVTADNVINASEAGNV--TVTGVLKN 706
QY 575 SPAEGPGPAAPASEVEGQKLRNVLVLSGEGY---IDFRIG-----DGED-----DET- 619
Db 707 VPADA-----ANTV-----VTVINGQTYTATVDSSTAGTWTVSVPGSDLTADADKTI 753
QY 620 -----EEGAGDMSQV 629
Db 754 DAKVTFTDAAGNSSSV 769

RESULT 15
US-07-757-022B-74
; Sequence 74, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1038 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-74

Query Match 3.4%; Score 118.5; DB 4; Length 1038;
Best Local Similarity 20.2%; Pred. No. 0.083;
Matches 67; Conservative 28; Mismatches 144; Indels 93; Gaps 12;

QY 10 CRPLVEKDPMTMKLWCAAGVNLGWRPNEDDAGNVKVPAPGRDPLTCDREGDG----- 61
Db 49 CKKYDKCCPDYESFCA-----EVKDNKKNRKTKKPTPKPPVVDGAGSLDNGDFKV 99
QY 62 -----EPKSAHTSPEKKKAK-----ELPMDATSSRVWILTS-----TLTTS 98
Db 100 TTPDSTTQHNVKSTSPKITTAKPINRPSLPNPSDTSKETS LTVNKETTIVETKETTTN 159
QY 99 KVVIIDANQPGTVVDQFTVCNAHVLCISSIPAAASDSDYPPGEMFLDSVNPEDPGADGVL 158
Db 160 KQTSTDGKEKTTSAKE-----TQSIKETSAXDLAPTS KVLAKPT----PKAETTT 205
QY 159 AGITLVGCATRCNVPRSNCSRGDTPVLDKQGGEV-----ATIANGKVNPSQSTEEAT 211
Db 206 KGPALT-----TPKEPTTTPKEPASTTPKEPTPTTIKSAPTTTPKEPAPTTT 252
QY 212 EA-----TEVPDPGPSEPE-----TATLRPGPLTEH-----VFTDPAPTPSSGPG 253
Db 253 KSAPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 312
QY 254 SENGPEPDSSSTRPEPEPSGDPGTGAGSSAAPT 285
Db 313 TPKEPAP-TTPKEPTTTPKEPAPTTTPKEPAPT 343

Search completed: August 23, 2004, 11:15:34
Job time : 24 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2004, 01:51:13 ; Search time 8246 Seconds  
(without alignments)  
13377.504 Million cell updates/sec

Title: US-10-019-495-26  
Perfect score: 3694  
Sequence: 1 ctggcagggttcctagtgcgc.....gtcgtataaaaaaaaaaaaaa 3694

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID          | Description        |
|------------|--------|-------------|--------|-------------|--------------------|
| 1          | 1478.8 | 40.0        | 2693   | 11 AK039130 | AK039130 Mus muscu |
| 2          | 844.8  | 22.9        | 870    | 13 BQ961808 | BQ961808 AGENCOURT |
| 3          | 782.2  | 21.2        | 1146   | 12 BM542648 | BM542648 AGENCOURT |
| 4          | 775.2  | 21.0        | 1201   | 13 BX460112 | BX460112 BX460112  |

|    |       |      |      |    |          |           |
|----|-------|------|------|----|----------|-----------|
| 5  | 772.2 | 20.9 | 1071 | 12 | BM547756 | AGENCOURT |
| 6  | 746.8 | 20.2 | 868  | 12 | BI906332 | 603063292 |
| 7  | 746.2 | 20.2 | 929  | 13 | BQ722520 | AGENCOURT |
| 8  | 721.4 | 19.5 | 901  | 13 | BX434805 | 603063292 |
| 9  | 717   | 19.4 | 981  | 13 | BX451957 | 603063292 |
| 10 | 713.4 | 19.3 | 748  | 12 | BI767677 | 603063292 |
| 11 | 683.8 | 18.5 | 690  | 14 | CD365632 | UI-H-FT2- |
| 12 | 675.8 | 18.3 | 884  | 14 | CD557812 | AGENCOURT |
| 13 | 665   | 18.0 | 721  | 12 | BI913764 | 603180726 |
| 14 | 664.2 | 18.0 | 727  | 12 | BI861206 | 603389971 |
| 15 | 660   | 17.9 | 862  | 13 | BU614626 | UI-M-EVO- |
| 16 | 642.2 | 17.4 | 1085 | 12 | BM803227 | AGENCOURT |
| 17 | 637.4 | 17.3 | 827  | 9  | AU079740 | AU079740  |
| 18 | 634.4 | 17.2 | 690  | 14 | CB852342 | UI-CF-FNO |
| 19 | 632.6 | 17.1 | 910  | 10 | BF969625 | 602271988 |
| 20 | 628   | 17.0 | 631  | 14 | CA422891 | UI-H-FLO- |
| 21 | 623   | 16.9 | 978  | 10 | BE797334 | 601587906 |
| 22 | 621.8 | 16.8 | 627  | 13 | BQ448063 | UI-H-EU1- |
| 23 | 618.8 | 16.8 | 622  | 13 | BX111494 | 601587906 |
| 24 | 618.8 | 16.8 | 1201 | 9  | AL577398 | AL577398  |
| 25 | 615.8 | 16.7 | 621  | 14 | CB216188 | NISC nq02 |
| 26 | 613   | 16.6 | 688  | 10 | BE907148 | 601499908 |
| 27 | 610.4 | 16.5 | 1027 | 13 | BQ071101 | AGENCOURT |
| 28 | 610   | 16.5 | 614  | 12 | BM666788 | UI-E-CL1- |
| 29 | 592.4 | 16.0 | 631  | 13 | BU633953 | UI-H-FL1- |
| 30 | 584.2 | 15.8 | 982  | 12 | BM449690 | AGENCOURT |
| 31 | 582.8 | 15.8 | 654  | 10 | BE795312 | 601586616 |
| 32 | 582.2 | 15.8 | 587  | 13 | BQ636226 | hd06b05.Y |
| 33 | 581.6 | 15.7 | 801  | 14 | CA319238 | UI-M-FW0- |
| 34 | 579.6 | 15.7 | 778  | 14 | CF536189 | UI-M-GI0- |
| 35 | 578.2 | 15.7 | 1013 | 13 | BX451956 | 603063292 |
| 36 | 576.6 | 15.6 | 771  | 14 | CF724148 | UI-M-GZ0- |
| 37 | 563.8 | 15.3 | 570  | 13 | BQ448060 | UI-H-EU1- |
| 38 | 561.4 | 15.2 | 571  | 14 | CA749006 | UI-H-FE1- |
| 39 | 558.6 | 15.1 | 875  | 13 | BU503257 | AGENCOURT |
| 40 | 558.4 | 15.1 | 770  | 10 | BE791544 | 601582616 |
| 41 | 558   | 15.1 | 783  | 14 | CB518578 | UI-M-GH0- |
| 42 | 557.6 | 15.1 | 748  | 14 | CF743414 | UI-M-GI0- |
| 43 | 555   | 15.0 | 1091 | 12 | BG294455 | 602391568 |
| 44 | 552.6 | 15.0 | 614  | 10 | BF952127 | QV1-NN022 |
| 45 | 548.6 | 14.9 | 765  | 13 | BQ572855 | UI-M-FD0- |

ALIGNMENTS

|            |  |                                    |   |                 |
|------------|--|------------------------------------|---|-----------------|
| RESULT 1   | AK039130   | 2693 bp                            | linear  | HTC 19-SEP-2003 |
| LOCUS      | Mus musculus adult male hypothalamus cDNA, RIKEN full-length   | enriched library, clone:A230101C01 | product:mitogen-activated protein kinase 8 interacting protein 3, full insert sequence. |                 |
| DEFINITION | AK039130   | GI:26333058                        |   |                 |
| ACCESSION  | AK039130   |                                    |   |                 |
| VERSION    | AK039130.1   |                                    |   |                 |
| KEYWORDS   | HTC; CAP trapper.  |                                    |   |                 |
| SOURCE     | Mus musculus (house mouse)   |                                    |   |                 |
| ORGANISM   | Mus musculus   |                                    |   |                 |
| REFERENCE  | 1  |                                    |   |                 |
| AUTHORS    | Carninci, P. and Hayashizaki, Y.   |                                    |   |                 |
| TITLE      | High-efficiency full-length cDNA cloning   |                                    |   |                 |
| JOURNAL    | Meth. Enzymol. 303, 19-44 (1999)   |                                    |   |                 |
| MEDLINE    | 99279253   |                                    |   |                 |
| PUBMED     | 10349636   |                                    |   |                 |
| REFERENCE  | 2  |                                    |   |                 |
| AUTHORS    | Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. |                                    |   |                 |
| TITLE      | Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes     |                                    |   |                 |
| JOURNAL    | Genome Res. 10 (10), 1617-1630 (2000)  |                                    |   |                 |
| MEDLINE    | 2049374  |                                    |   |                 |
| PUBMED     | 11042159   |                                    |   |                 |

|           |   |  |
|-----------|---|--|
| REFERENCE | 3   | Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer |
| JOURNAL   | Genome Res. 10 (11), 1757-1771 (2000)   |  |
| MEDLINE   | 20530913  |  |
| PUBMED    | 11076861  |  |
| REFERENCE | 4   |  |
| AUTHORS   | The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  |  |
| TITLE     | Functional annotation of a full-length mouse cDNA collection  |  |
| JOURNAL   | Nature 409, 685-690 (2001)  |  |
| REFERENCE | 5   |  |
| AUTHORS   | The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  |  |
| TITLE     | Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  |  |
| JOURNAL   | Nature 420, 563-573 (2002)  |  |
| REFERENCE | 6 (bases 1 to 2693)   |  |
| AUTHORS   | Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. |  |
| TITLE     | Direct Submission   |  |
| JOURNAL   | Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)   |  |
| COMMENT   | cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.go.jp/ URL:http://fantom.gsc.riken.go.jp/.  |  |
| FEATURES  | Location/Qualifiers   |  |
| source    | 1..2693   |  |
|           | /organism="Mus musculus"  |  |
|           | /mol_type="mRNA"  |  |
|           | /strain="C57BL/6J"  |  |
|           | /db_xref="FANTOM_DB:A230101C01"   |  |
|           | /db_xref="MGI:2403586"  |  |
|           | /db_xref="taxon:10090"  |  |
|           | /clone="A230101C01"   |  |
|           | /sex="male"   |  |
|           | /tissue type="hypothalamus"   |  |
|           | /clone_lib="RIKEN full-length enriched mouse cDNA library"  |  |
|           | /dev_stage="adult"  |  |
|           | <1..1764  |  |
| CDS       | /note="unnamed protein product; mitogen-activated protein kinase 8 interacting protein 3 (MGD MGI:1353598, GB NM_013931, evidence: BLASTN, 99%, match=2586) putative"   |  |
|           | /codon_start=1  |  |
|           | /protein_id="BAC30247.1"  |  |
|           | /translation="STHPSPEKKKAKETPEADATSSRVWILTSTLTTSKVVIIDANQP GTIVDQFTVCNAHVLCISSIPAAASDSYPPGEMFLDSVDNPDSDGAGVLGITLVGC ATRCNVPRNSCSRSGDTPLVDKGQDVAATTANGKNVPSQSTEEATEATEVDPGPSES EATTVRPGPLTEHVFTDPAFTPSSSTOPASENGSENGTIVQPQVEPSGELSTTTSSA APTMWLGAQNGWLYVHSAVANWKKCLHSIKLXDSLVLSHVHVKRVLVALADGTLAIFH RGEDGQWDLSNYHLMDLGHPHHSIRCMVNDRWVCYKKNKHVIQPKTMQIEKSFDA HPRRESQVROLAWIGDGVVVSIRLDSITRLRYHAHTHQHLDVIEPYVSKMLGTGKLG FFSFVRITALLTAGNRLMWGTNGVVISIPLTETVVLHRGQLLGRANKTSPTSGETIR PGGIHHVYDDSDSKAASSFIPYCSMAQAOLCFHGRDVKFFVSPGNVLATLNGSV LDSPSEGPGPAPAAADAEGQKLNALVLSGGEGYIDFRIGDGEDDETEECAGDVNQTK PSLSKAERSHIIVWQVSYTPE"  |  |
|           | polyA_signal  | 2679..2684   |
|           | /note="putative"  |  |
|           | polyA_site  | 2693   |
|           | /note="putative"  |  |
| ORIGIN    |   |  |
|           | Query Match   | 40.0%; Score 1478.8; DB 11; Length 2693;   |
|           | Best Local Similarity   | 75.3%; Pred. No. 1.7e-234;   |
|           | Matches 1970; Conservative  | 0; Mismatches 612; Indels 34; Gaps 9;  |
| QY        | 424   | AGCGCCACACGCTCTCCGGAAGAAGAGGCAAGGAGTCCCTGAAATGGAGCGCCACC 483   |
| Db        | 1   | AGCACACACCCATCACCTGAGAAGAAGAGGCAAGGAAACCCCTGAGGACATGCTACC 60   |
| QY        | 484   | TCCAGCGGGTGTGGATCCTGACACGACCCCTGACCAACAGGAAGGTGGTGCATCGAC 543  |
| Db        | 61  | TCCAGTCGGGTATGGATCCTCACCAGCACCCCTGACCAACAGCAAGGTGGTGCATGAT 120   |
| QY        | 544   | GCCAAACAGCGGGCACGGTGGTGGACCAAGTTCACCGTCTGCAACCGGCACGCTGTGC 603   |
| Db        | 121   | GCCAAACAGCGGACCAATTTGGGATCAGTTTCAAGTCTGCAATGCCCCACGTCCTGTGT 180  |
| QY        | 604   | ATCTCCAGCATCCCGCGGCGACGACGCGGACTACCTCCCGGGGAGATGTTCTCTGGAC 663   |
| Db        | 181   | ATCTCCAGCATCTCTCGGCGCAGTGACAGTGACTATCCCCCTGGGGAGATGTTCTCTAGAC 240  |
| QY        | 664   | AGCGACGTGAACCCAGAGGACCGCGGCGCAGATGGCGTGTGCGCGGTATCACCTGGTGTG 723   |
| Db        | 241   | AGTGATGTGAACCTGAAGATTTCAGGTGCTGATGGTGTGTGCTGGCTGGCATCACCGTGTG 300  |
| QY        | 724   | GGCTGTGCCACCGCTGCAACGTCGCGGGAGCAACTGCTCTCTCCCGAGGGGACACCCCA 783  |
| Db        | 301   | GGGTGTCTACCGCTGCAATGTTCCACGTAGCAACTGTTCTCCTCAGGAGGACACCCCA 360   |
| QY        | 784   | GTGCTAGACAAGGGGACGGGGAGGTGGCCACCATCGCCAAACGGGAAGGTCAACCGCTCC 843   |
| Db        | 361   | GTAATGGACAAGGGGACGGGGAGTGTGGCGACCACTGCTGCAATGGGAAGGTCAACCGCTCC 420   |
| QY        | 844   | CAGTCCACAGAGGAGGCCACAGAGGCCACGGAGTGCCAGACCCCTGGGCCCCAGCGAGCCA 903  |
| Db        | 421   | CAATCCACAGAGAAGCCACAGAAGCCACGGAGTGCCAGACCCCTGGTCCCAGCGAGTCA 480  |
| QY        | 904   | GAGACAGCCACATTGCGGCCCGCGCTCTTCACAGAGCAGTCTTCACTGACCCAGCCCCG 963  |
| Db        | 481   | GAAGCAACGACAGTCCGGCCCGCGCTCTTCACAGAGCATGTCTTTACTGACCCGACCCC 540  |
| QY        | 964   | ACCCCGTCTCTGGCCCCCAGCCTGGCAGCGAGAACGGGCCAGAGCCTGACAGCAGCAGC 1023   |
| Db        | 541   | ACCCCATCTCTCCAGCACCCAGCCTGCGAGTGAGAAATGGGTTCAGAGTCCAATGGCACCAT 600   |
| QY        | 1024  | ACACGGCCAGAGCCAGCCAGCGGGGACCCCAACGGGAGCAGGCGAGCAGTGTGACCCC 1083  |
| Db        | 601   | GTACAGCCTCAGGTGGAGGCCAGTGGGGAACCTCTCAACAACAACCAAGTAGCGCTGCACCC 660   |
| QY        | 1084  | ACCATGTGGCTGGAGGCCCAGAACCGGTGGCTCTATGTGTCACTCGGCTGTGGCCAACTGG 1143   |
| Db        | 661   | ACTATGTGGCTAGAGGCCAGAAATGGCTGGCTCTATGTGCATTCAGCGGTAGCCAACTGG 720   |
| QY        | 1144  | AAGAAGTGCCTGCACCTCCATCAAGCTGAAGGATTTCTGTGCTGAGCCTGGTGCATGTCAA 1203   |
| Db        | 721   | AAGAAGTGTCTGTCACTCCATCAAGCTAAAGAAGCTCTGTGCTGAGCCCTGGTGCATGTCAA 780   |



cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM2567 row: 1 column: 21  
High quality sequence stop: 728.  
Location/Qualifiers  
1. .870  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6379940"  
/tissue\_type="epithelioid carcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_42"  
/note="Organ: pancreas; Vector: pOTB7; Site 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library. |"

Db 608 AGACATCCCCACCTCTGGGAGGGCGCCGTCCTCCGGGGGCATCATCCACGTGTATGGCG 667  
QY 1802 ATGACAGAGTACAGGGCGGCCAGCAGCTTCATCCCTACTGCTCCATGGCCCCAGGCC 1861  
Db 668 ATGACAGAGTACAGGGCGGCCAGCAGCTTCATCCCTACTGCTCCATGGCCCCAGGCC 727  
QY 1862 AGCTATGCTTCCATGGGCACCGCATGCCGTGAAGTTCCTTGTCTCGGTGCCAGGGAACG 1921  
Db 728 AGCTATGCTTCCATGGGCACCGCATGCCGTGAAGTTCCTTGTCTCGGTGCCAGGGAACG 787  
QY 1922 TGCTGGCCACCCTGAATGGCAGTGTGCTGGACAGCCAGCCAGGGCCCCCTGGCCAGCTG 1981  
Db 788 TGCTGGCCACCCTGAATGGCAGTGTGCTGGACAGCCAGCCAGGGCCCCCTGGCCAGCTG 847  
QY 1982 CCCCT-GCCTCGGAGGTCGAGG 2002  
Db 848 CCCTTGGCCTCGGAGGTCGAGG 869

RESULT 3  
BM542648  
LOCUS BM542648  
DEFINITION BM542648.1 GI:18772346  
ACCESSION BM542648  
VERSION BM542648.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1146)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM2185 row: h column: 17  
High quality sequence stop: 634.

FEATURES  
source  
1. .1146  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5520832"  
/tissue\_type="leiomyosarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_71"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: Sall; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2.1 kb."

ORIGIN  
Query Match 21.2%; Score 782.2; DB 12; Length 1146;  
Best Local Similarity 95.0%; Pred. No. 2.8e-119;  
Matches 852; Conservative 0; Mismatches 38; Indels 7; Gaps 4;  
QY 1704 CCTGCACCGAGGCCAGCTCCTGGGCTCCGAGCCATAAGACATCCCCACCTCTGGGGA 1763  
Db 8 CGTGGCCGAGGCCAGCTCCTGGGCTCCGAGCCATAAGACATCCCCACCTCTGGGGA 67  
QY 1764 GGGCGCCCGTCCCGGGGCATCATCCAGTGTATGGCGATGACAGCAGTGACAGGGCGGC 1823  
Db 68 GGGCGCCCGTCCCGGGGCATCATCCAGTGTATGGCGATGACAGCAGTGACAGGGCGGC 127  
QY 1824 CAGCAGTTCATCCCTACTGTCTCCATGGCCCCAGGCCAGCTATGCTTCCATGGCACCG 1883

Query Match 22.9%; Score 844.8; DB 13; Length 870;  
Best Local Similarity 99.4%; Pred. No. 1.2e-129;  
Matches 857; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
QY 1142 GGAAGAAGTGCCTGCACCTCCATCAAGCTGAAGATTCTGTGCTGAGCCTGGTGTGATGTC 1201  
Db 8 GGAAGAAGTGCCTGCACCTCCATCAAGCTGAAGATTCTGTGCTGAGCCTGGTGTGATGTC 67  
QY 1202 AAGCCGTGTGCTGGTGGCTCTGGCGGACGGACCTGGCCATCTCCACCGTGGTGAAG 1261  
Db 68 AAGCCGTGTGCTGGTGGCTCTGGCGGACGGACCTGGCCATCTCCACCGTGGTGAAG 127  
QY 1262 ATGGCCAGTGGGATCTGAGCAACTATCACTAATGACCTGGGCGACCCGACCACTCCA 1321  
Db 128 ATGGCCAGTGGGATCTGAGCAACTATCACTAATGACCTGGGCGACCCGACCACTCCA 187  
QY 1322 TCCGTGTCATGGCTGTGTGTACGACCGCGTGTGGTGGCTACAGAACAAAGTGCACG 1381  
Db 188 TCCGTGTCATGGCTGTGTGTACGACCGCGTGTGGTGGCTACAGAACAAAGTGCACG 247  
QY 1382 TCATCCAGCCCAAGACCATGCAGATAGAGAAGTCAATTGACGCCACCCCGCGGGGAGA 1441  
Db 248 TCATCCAGCCCAAGACCATGCAGATAGAGAAGTCAATTGACGCCACCCCGCGGGGAGA 307  
QY 1442 GCCAGTGCAGCTGCGGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCTGGACT 1501  
Db 308 GCCAGTGCAGCTGCGGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCTGGACT 367  
QY 1502 CCACCCCTGAGGCTCTACCATGCACACGACCGACCATCTACAGGACGTGGACATTGAGC 1561  
Db 368 CCACCCCTGAGGCTCTACCATGCACACGACCGACCATCTACAGGACGTGGACATTGAGC 427  
QY 1562 CCTACGTGAGCAAGATGCTAGGCACCTGGCAAGCTGGGTTCTCCTTCGTACGCATCACGG 1621  
Db 428 CCTACGTGAGCAAGATGCTAGGCACCTGGCAAGCTGGGTTCTCCTTCGTACGCATCACGG 487  
QY 1622 CCCTGCTTGTTCGGGGAGCCGGCTCTGGGTGGGACCGGCAACGGAGTGGTCTATCTCA 1681  
Db 488 CCCTGCTTGTTCGGGGAGCCGGCTCTGGGTGGGACCGGCAACGGAGTGGTCTATCTCA 547  
QY 1682 TCCCCCTGACAGAGACTGTGGTCTGTGACCGAGGCCAGCTCCTGGGGCTCCGAGCCATA 1741  
Db 548 TCCCCCTGACAGAGACTGTGGTCTGTGACCGAGGCCAGCTCCTGNGGCTCCGAGCCATA 607  
QY 1742 AGACATCCCCACCTCTGGGAGGGCGCCGTCCTCCGGGGCATCATCCAGTGTATGGCG 1801

```
|||||
Db 128 CAGAGCTTCCATCCCTACTGCTCCATGGCCAGGCCAGCTATGCTTCCATGGCACCG 187
QY 1884 CGATCCCGTGAAGTTCTTTGTCTCGGTGCCAGGGAACGTGCTGGCCACCTGAAATGGCAG 1943
Db 188 CGATCCCGTGAAGTTCTTTGTCTCGGTGCCAGGGAACGTGCTGGCCACCTGAAATGGCAG 247
QY 1944 TGTGCTGGACAGCCAGCCAGGCGCCTGGGCCAGCTGCCCTGCTCGGAGTCCAGGG 2003
Db 248 TGTGCTGGACAGCCAGCCAGGCGCCTGGGCCAGCTGCCCTGCTCGGAGTCCAGGG 307
QY 2004 CCAGAACTGCGGAACAGTGTGCTGTGCTGAGCGCGGGGAGGCTFACATCGACTTCCGCGAT 2063
Db 308 CCAGAACTGCGGAACAGTGTGCTGTGCTGAGCGCGGGGAGGCTFACATCGACTTCCGCGAT 367
QY 2064 TGGAGACGGAGAGGACGACGAGACGGAGGAGGCGCAGGGACATGAGCCAGGTGAAGCC 2123
Db 368 TGGAGACGGAGAGGACGACGAGACGGAGGAGGCGCAGGGACATGAGCCAGGTGAAGCC 427
QY 2124 CGTGCTGTCCAAAGCAGAGCGGAGTCAATCATCGTGTGGCAGGTGTCTTACACCCCGA 2183
Db 428 CGTGCTGTCCAAAGCAGAGCGGAGTCAATCATCGTGTGGCAGGTGTCTTACACCCCGA 487
QY 2184 GTGAAGCTGCTGCCCTGCTGCCCGCAGCTGCTFACATAGGACCCCGACCTGACCCCG 2243
Db 488 GTGAAGCTGCTGCCCTGCTGCCCGCAGCTGCTFACATAGGACCCCGACCTGACCCCG 547
QY 2244 GCCCGGCCCGGGGTAGCCAGCCAGGCGCGCCCGCCCTCTTCTAACTCTCAACCTGC 2303
Db 548 GCCCGGCCCGGGGTAGCCAGCCAGGCGCGCCCGCCCTCTTCTAACTCTCAACCTGC 607
QY 2304 AGCTTTCACTGAGTCTGGCCCTCCAGCGGCGAGGAGTGGGGATGCGGATCAGCTG 2363
Db 608 AGCTTTCACTGAGTCTGGCCCTCCAGCGGCGAGGAGTGGGGATGCGGATCAACTG 667
QY 2364 GGAGGAGGAGGGAGGGGTGCTTCCACCCGAGGGAAGATGCTCTCGGACAGTTTCCCG 2423
Db 668 GGAGGAGGAGGGGA -GGGTGCTTCCACCCGAGGGAAGATGCTCCCGAGACAGTTTCCCG 726
QY 2424 GGCAGCTCTGGCCAGCTTCCAGCCAGCTCCTCAAGTCCAGGCGACCTTGGGCCAGC 2483
Db 727 GGCAGCTCTGGCCAGCTTCCAGCCAGAGCCCTCAAGT -CAGGGACCTTGGGCCAGC 785
QY 2484 GCAGGC -AGAAATCCGAGGTGCTGCTGCTACCTGGGCTCTACTCTCCCGACACCC 2542
Db 786 GCAGCCAAAATCCAAAGGGGCCCGGCTTACCTGGGCTCTACTCTCCCGAAACCC 845
QY 2543 T----GGAGGAGGCGAGGGCTCCCGCGCGGAGGCTGCTGCCCTGGGCCACCTC 2595
Db 846 TTGGAGGAAGCCCGAGCTTCTCCCGCCCGCAAGACATCTCCCTGGCGCCCCC 902
```

```
RESULT 4
BX460112
LOCUS
DEFINITION BX460112 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
CS0DF011YK03 5-PRIME, mRNA sequence.
ACCESSION BX460112
VERSION BX460112.1 GI:31023201
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
```

Invitrogen. This sequence belongs to sequence cluster 8197.f For more information about this cluster, see <http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DF011AF02QP1&cluster=8197.f>. Contact : Feng Liang Email : [fliang@lifetech.com](mailto:fliang@lifetech.com) URL : <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DF011AF02QP1.

Location/Qualifiers  
1. 1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DF011YK03"  
/tissue\_type="FETAL BRAIN"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL BRAIN"  
/note="Organ: brain; Vector: pCMVSPORT\_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

```
Query Match 21.0%; Score 775.2; DB 13; Length 1201;
Best Local Similarity 92.2%; Pred. No. 4.1e-118;
Matches 807; Conservative 24; Mismatches 39; Indels 5; Gaps 3;

QY 2072 GAGAGGACGACGAGACGAGGAGGCGCAGGGACATGAGCAGGTGAAGCCCGTGTGT 2131
Db 58 GANAGGACGACGAGACGAGGAGGCGCAGGGACATGAGCAGGTGAAGCCCGTGTGT 117
QY 2132 CCAAGGACGAGCGCAGTCAATCATCGTGTGGCAGGTGTCTTACCTCTCAACCTGAGTGAAGCT 2191
Db 118 CCAAGGACGAGCGCAGTCAATCATCGTGTGGCAGGTGTCTTACCTCTCAACCTGAGTGAAGCT 177
QY 2192 GCTGCCCTGCTGGCCCGACCTGTACATAGGACCCCGACCTGACCCCGCCCGGCC 2251
Db 178 GCTGCCCTGCTGGCCCGACCTGTACATAGNNNBCCCGACCTGABCCCGCCCGGCC 237
QY 2252 CGCGGGTAGCCAGCCAGCGCGCGCCCGCCCTCTTCTAACCTCTCAACCTGAGTTCATCA 2311
Db 238 CGCGGGTAGCCAGCCAGCGCGCGCCCGCCCTCTTCTAACCTCTCAACCTGAGTTCATCA 297
QY 2312 CCTGAGTCTGGCCCTCCAGCGGGCAGGAGTCCGGGATGCGGATCAGTGGGAGGAGG 2371
Db 298 CCTGAGTCTGGCCCTCCAGCGGGCAGGAGTCCGGGATGCGGATCAGTGGGATGAGG 357
QY 2372 AGGGAGGGGTGCTTCCACCCGAGGGAAGATGCTCTCGGACAGTTTCCCGGCAGCTC 2431
Db 358 ANGNANNTNCTTCCACCCGAGGGAAGATGCTCTCGGACAGTTTCCCGGCAGCTC 417
QY 2432 CTGGCCAGCTTCCAGCCAGAGTCTCAAGTCCAGGACACCTTGGGCCAGCGCAGGAG 2491
Db 418 CTGGCCANCTTCCAGCCAGAGTCTCAAGTCCAGGACACCTTGGGCCAGCGCAGGAG 477
QY 2492 AATCCGAGGTGCTCTGCTCTACCTGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2551
Db 478 AATCCGAGGTGCTCTGCTCTACCTGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 537
QY 2552 CAGGGGCTCCCGCGCCGAGGCTGCTGCCCTGGGCCACCTCTGCTGCTGCTGCTGCTGCT 2611
Db 538 CAGGGGCTCCCGCGCCGAGGCTGCTGCCCTGGGCCACCTCTGCTGCTGCTGCTGCTGCTGCT 597
QY 2612 GGCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2671
Db 598 GGCCA-NCTGNATNCTGGNNCTNAATCTGNCTAGGGAGCTGGGCCAGGCACCTAGCCTT 656
QY 2672 TGCCAGGGAGGTGGGCTCAGGCTGCCAGGTGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2731
Db 657 TGCCAGGGAGGTGGGCTCAGGCTGCCAGGTGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 716
QY 2732 GGCTCCCGCTGCTCAAGCTCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2791
```

|  |      |   |  |                       |     |
|--|------|---|--|-----------------------|-----|
| Db   | 717  | GGCCTCCCGTCGTC  | AAAGCCTCTATCCTGTCTGTCTCCCA                   | CCCCCAGCTGTTCCCTGCCAG | 776 |
| QY   | 2792 | GGAGCTGGCATAAAAG  | CACGAGGCCCGGCTCCCTGGGGCAGCTGCTTGAGAACAGAGACT | 2851                  |     |
| Db   | 777  | GGAGCTGGCATAAAAG  | CAMGAGSCCGGCTCYCTGGGCAGCTKCTTGAGAACARAGACT   | 836                   |     |
| QY   | 2852 | GCTACCCCATCTGCCC  | ATGCAGGCAGGCTCTTGGCAGCCCCGTTCTGACCCCGTGTCCCC | 2911                  |     |
| Db   | 837  | GYTACCCCATCTGSCC  | ATGCAGSMAGCYVTTTSCMACCCCGTTCTGVACSGK---      | CCC                   | 893 |
| QY   | 2912 | CCAGGCTCTGCTGGG   | CAGAGACTCACCTTGGAG                           | 2946                  |     |
| Db   | 894  | CCARGYKCTG-CTGGG  | GAGAAAAATCACYTTGSAG                          | 927                   |     |
| RESULT 5   |      |   |  |                       |     |
| BM547756   |      |   |  |                       |     |
| LOCUS  |      |   |  |                       |     |
| DEFINITION   |      |   |  |                       |     |
| AGENCOURT_6507064 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5727893  |      |   |  |                       |     |
| 5', mRNA sequence.   |      |   |  |                       |     |
| BM547756   |      |   |  |                       |     |
| BM547756.1 GI:18781796   |      |   |  |                       |     |
| EST.   |      |   |  |                       |     |
| Homo sapiens (human)   |      |   |  |                       |     |
| ORGANISM   |      |   |  |                       |     |
| Homo sapiens   |      |   |  |                       |     |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.   |      |   |  |                       |     |
| REFERENCE  |      |   |  |                       |     |
| 1 (bases 1 to 1071)  |      |   |  |                       |     |
| NIH-MGC http://mgs.nci.nih.gov/.   |      |   |  |                       |     |
| National Institutes of Health, Mammalian Gene Collection (MGC)   |      |   |  |                       |     |
| JOURNAL  |      |   |  |                       |     |
| Unpublished (1999)   |      |   |  |                       |     |
| COMMENT  |      |   |  |                       |     |
| Contact: Robert Strausberg, Ph.D.<br>Email: cgapbs-r@mail.nih.gov<br>Tissue Procurement: Invitrogen<br>cDNA Library Preparation: Life Technologies, Inc.<br>cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)<br>DNA Sequencing by: Agencourt Bioscience Corporation<br>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov<br>Plate: LLAM12722 row: 1 column: 06<br>High quality sequence stop: 676. |      |   |  |                       |     |
| FEATURES   |      |   |  |                       |     |
| source   |      |   |  |                       |     |
| 1..1071  |      |   |  |                       |     |
| /organism="Homo sapiens"   |      |   |  |                       |     |
| /mol_type="mRNA"   |      |   |  |                       |     |
| /db_xref="taxon:9606"  |      |   |  |                       |     |
| /clone="IMAGE:5727893"   |      |   |  |                       |     |
| /tissue_type="hippocampus"   |      |   |  |                       |     |
| /lab_host="DH10B"  |      |   |  |                       |     |
| /clone_lib="NIH_MGC_124"   |      |   |  |                       |     |
| /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: EcoRV (destroyed); Site 2: NotI; RNA source male hippocampus, age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 0.9-4 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 012."   |      |   |  |                       |     |
| ORIGIN   |      |   |  |                       |     |
| Query Match  |      |   |  |                       |     |
| Best Local Similarity  |      |   |  |                       |     |
| Matches 805; Conservative 0; Mismatches 18; Indels 2; Gaps 2;  |      |   |  |                       |     |
| QY   | 666  | CGACGTGAACCCAGAGACCCGGCGCAGATGGCGTATCACCCCTGGTGGG         | 725  |                       |     |
| Db   | 1    | CGACGTGAACCCAGAGACCCGGCGCAGATGGCGTATCACCCCTGGTGGG         | 60   |                       |     |
| QY   | 726  | CTGTGCCACCCGCTGCAACGTGCCGGAGCAACTGCTCTCCCGAGGGGACACCCCACT | 785  |                       |     |
| Db   | 61   | CTGTGCCACCCGCTGCAACGTGCCGGAGCAACTGCTCTCCCGAGGGGACACCCCACT | 120  |                       |     |

|  |      |   |      |
|--|------|---|------|
| Qy   | 786  | GCTAGACAAGGGGCGAGGGGAGGTGGCCACCATCGCCACGGGAAGGTCAACCCCGTCCCA  | 845  |
| Db   | 121  | GCTAGACAAGGGGCGAGGGGAGGTGGCCACCATCGCCACGGGAAGGTCAACCCCGTCCCA  | 180  |
| Qy   | 846  | GTCCACAGAGGAGGCCACAGAGGCCACCGAGGTGCCAGACCCTGGGCCCGAGGCCAGA    | 905  |
| Db   | 181  | GTCCACAGAGGAGGCCACAGAGGCCACCGAGGTGCCAGACCCTGGGCCCGAGGCCAGA    | 240  |
| Qy   | 906  | GACAGCCACATTGGGCCCCCGGCTCTCTACAGAGCACGTTCTTCACTGACCCAGCCCGGAC | 965  |
| Db   | 241  | GACAGCCACATTGGGCCCCCGGCTCTCTACAGAGCACGTTCTTCACTGACCCAGCCCGGAC | 300  |
| Qy   | 966  | CCCGTCTCTGGCCCCCGGCTGGCAGCGGAGAACCGGCCAGAGCCTGACAGCAGCAGCAC   | 1025 |
| Db   | 301  | CCCGTCTCTGGCCCCCGGCTGGCAGCGGAGAACCGGCCAGAGCCTGACAGCAGCAGCAC   | 360  |
| Qy   | 1026 | ACGGCCAGAGCCAGAGCCAGCGGGGACCCACGGGAGCAGGCAGCAGTGTGACCCAC      | 1085 |
| Db   | 361  | ACGGCCAGAGCCAGAGCCAGCGGGGACCCACGGGAGCAGGCAGCAGTGTGACCCAC      | 420  |
| Qy   | 1086 | CATGTGGCTGGGAGCCCGAGACGGCTGGCTCTATGTGCATCTGGCTGTGGCCAACTGGAA  | 1145 |
| Db   | 421  | CATGTGGCTGGGAGCCCGAGACGGCTGGCTCTATGTGCATCTGGCTGTGGCCAACTGGAA  | 480  |
| Qy   | 1146 | GAAGTGCCTGCATCCATCAAGCTGAAGGATCTGTGCTGAGCCTGGTGCATGTCAAAGG    | 1205 |
| Db   | 481  | GAAGTGCCTGCATCCATCAAGCTGAAGGATCTGTGCTGAGCCTGGTGCATGTCAAAGG    | 540  |
| Qy   | 1206 | CCGTGTGCTGGTGGCTCTGGCGGACGGGACCTGGCCATCTTCCACCGTGGTGAAGATGG   | 1265 |
| Db   | 541  | CCGTGTGCTGGTGGCTCTGGCGGACGGGACCTGGCCATCTTCCACCGTGGTGAAGATGG   | 600  |
| Qy   | 1266 | CCAGTGGGATCTGAGCAACTATACCTAATGGACCTGGGCCACCCGACCATCCATCCG     | 1325 |
| Db   | 601  | CCAGTGGGATCTGAGCAACTATACCTAATGGACCTGGGCCACCCGACCATCCATCCG     | 660  |
| Qy   | 1326 | CTGCATGGCTGTGTGTACGACCGCGTGTGGTGGCTACAGAACAAAGGTGCACGTCA      | 1385 |
| Db   | 661  | CTGCATGGCTGTGTGTACGACCGCGTGTGGTGGCTACAGAACAAAGGTGCACGTCA      | 720  |
| Qy   | 1386 | CCAGCCCAAGACCATGCAGATAGAGAAATCATTTGAAGCCCCCCCCCGGTGGGAGAGCC   | 1444 |
| Db   | 721  | CCAGCCCAAGACCATGCAGATAGAGAAATCATTTGAAGCCCCCCCCCGGTGGGAGAGCC   | 780  |
| Qy   | 1445 | AGGTGCGGCGAGTGGCGTGG-ATCGCGCATGGCGTATGGTGTCC                  | 1488 |
| Db   | 781  | CAGTGGCGCAACTGGGTGGAATCGGCAAGGGCTTATGGTGTCC                   | 825  |
| RESULT 6   |      |   |      |
| BI906332   |      |   |      |
| LOCUS  |      |   |      |
| DEFINITION   |      |   |      |
| BI906332 603063292F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5212463 5', mRNA sequence.  |      |   |      |
| ACCESSION  |      |   |      |
| BI906332   |      |   |      |
| VERSION  |      |   |      |
| BI906332.1 GI:16169009   |      |   |      |
| KEYWORDS   |      |   |      |
| EST.   |      |   |      |
| SOURCE   |      |   |      |
| ORGANISM   |      |   |      |
| Homo sapiens (human)   |      |   |      |
| Homo sapiens   |      |   |      |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.   |      |   |      |
| REFERENCE  |      |   |      |
| 1 (bases 1 to 868)   |      |   |      |
| NIH-MGC http://mgs.nci.nih.gov/.   |      |   |      |
| National Institutes of Health, Mammalian Gene Collection (MGC)   |      |   |      |
| JOURNAL  |      |   |      |
| Unpublished (1999)   |      |   |      |
| COMMENT  |      |   |      |
| Contact: Robert Strausberg, Ph.D.<br>Email: cgapbs-r@mail.nih.gov<br>Tissue Procurement: Life Technologies, Inc.<br>cDNA Library Preparation: Life Technologies, Inc.<br>cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)<br>DNA Sequencing by: Incyte Genomics, Inc.<br>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: |      |   |      |

http://image.llnl.gov  
Plate: LLAM1533 row: g column: 24  
High quality sequence stop: 812.  
Location/Qualifiers  
1. 868  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5212463"  
/tissue\_type="leukocyte"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_118"  
/note="Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV  
(destroyed); RNA source leukocytes from anonymous pool of  
non-activated adult donors. Library is oligo-dT primed  
and directionally cloned (EcoRV site is destroyed upon  
cloning). Average insert size 1.7 kb, insert size range  
1.2-3.3 kb. Library is normalized and enriched for  
full-length clones and was constructed by C. Gruber  
(Invitrogen). Research Genetics tracking code 027. Note:  
this is a NIH\_MGC Library."

ORIGIN

Query Match 20.2%; Score 746.8; DB 12; Length 868;  
Best Local Similarity 97.2%; Pred. No. 1.9e-113;  
Matches 824; Conservative 0; Mismatches 17; Indels 7; Gaps 6;  
QY 1096 GGAGCCAGAACGGCTGCTTATGTGCATCGGCTGTGGCCAACTGGAAGAGTGCTG 1155  
Db 1 GGAGCCAGAACGGCTGCTTATGTGCATCGGCTGTGGCCAACTGGAAGAGTGCTG 60  
QY 1156 CACTCCA-TCAAGCTGAAGATCTGTGCTGAGCCTGGTGATGTCAAAGGCGGTGCT 1214  
Db 61 CACTCCAGTCAAGCTGAAGATCTGTGCTGAGCCTGGTGATGTCAAAGGCGGTGCT 120  
QY 1215 GGTGGCTCTGGCGGACGGGACCTGGCCATCTTCCACCGTGGTGAAGATGCGCAGTGGGA 1274  
Db 121 GGTGGCTCTGGCGGACGGGACCTGGCCATCTTCCACCGTGGTGAAGATGCGCAGTGGGA 180  
QY 1275 TCTGAGCAACTATCACCTAATGGACCTGGGACCCCGCACCACTCCATCCGCTGCATGGC 1334  
Db 181 TCTGAGCAACTATCACCTAATGGACCTGGGACCCCGCACCACTCCATCCGCTGCATGGC 240  
QY 1335 TGTGTGTACGACCGCGTGTGGTGTGGCTACAAGAACAAAGTGCACGTCACTCCAGCCCAA 1394  
Db 241 TGTGTGTACGACCGCGTGTGGTGTGGCTACAAGAACAAAGTGCACGTCACTCCAGCCCAA 300  
QY 1395 GACCATGCAGATAGAGAAGTCATTTGACGCCACCCCGCGGGAGAGCCAGGTGCGGCA 1454  
Db 301 GACCATGCAGATAGAGAAGTCATTTGACGCCACCCCGCGGGAGAGCCAGGTGCGGCA 360  
QY 1455 GCTGGCTGGATCGGCGATGGCGTATGGTGTCCATCCGCTGGACTCCACCCCTGAGGCT 1514  
Db 361 GCTGGCTGGATCGGCGATGGCGTATGGTGTCCATCCGCTGGACTCCACCCCTGAGGCT 420  
QY 1515 CTACCATGCACACACGACCCAGCATCTACAGGACGTGGACATTGAGCCCTACGTCAGCAA 1574  
Db 421 CTACCATGCACACACGACCCAGCATCTACAGGACGTGGACATTGAGCCCTACGTCAGCAA 480  
QY 1575 GATGTAGGCACCTGGCAAGCTGGGTTTCTCTTCGTACGCATCACGGCCCTGCTTTCGC 1634  
Db 481 GATGTAGGCACCTGGCAAGCTGGGTTTCTCTTCGTACGCATCACGG-CCTGCTTTCGC 539  
QY 1635 GGGCAGCGGCTCTGGGTGGGACCGGCAACGGAGTGTCTCTCCATCCCCCTGACAGA 1694  
Db 540 GGGCAGCGGCTCTGGGTGGGACCGGCAACGGAGTGTCTCTCCATCCCCCTGACAGA 599  
QY 1695 GACTGTGGTCTGCACCGAGGCCAGCTCCTTGGGGCTCCGAGCCCAATAGACATCCCCCAC 1754  
Db 600 GACTGTGGTCTGCACCGAGGCCAGCTCCTTGGGGCTCCGAGCCCAATAGACATCCCCCAC 659  
QY 1755 CTCTGGGAGGGCGCCGCTCCCGGGGGCATCATCCACGTGTATGGCGATGACAGAGTGA 1814

Db 660 CTCTGGGAGGGCGCCCGTCCCGGGGCATCATCCACGTGTATGGCGATGACAGCAGTGA 719  
QY 1815 CAGGGCGGCCAGCAGCTTTCATCCCTACTGCTCCATGGCCCGCCAGGCCAGCTATGCTTCCA 1874  
Db 720 CAGGGCGG-CAGCAGCTTCAT-CCCTACTGCTCCATGGCC--AGCCAGCTATGCTTCCA 775  
QY 1875 TGGGCACCGCGATGCGCGTG-AAGTCTTTTGTCTCGGTGCCAGGGAACGTGCTGGCCACCC 1933  
Db 776 TGGGCACCGCGATGCGCGTGAAGTCTTTGTCTCGGTGCCACGCGCCGCTGCTTGCACC 835  
QY 1934 TGAATGGC 1941  
Db 836 TGAATGTC 843  
RESULT 7  
BQ722520  
LOCUS BQ722520 929 bp mRNA linear EST 16-JUL-2002  
DEFINITION AGENCOURT 8305079 Lupski\_sympathetic\_trunk Homo sapiens cDNA clone  
IMAGE:6193163 5', mRNA sequence.  
ACCESSION BQ722520  
VERSION BQ722520.1 GI:21861417  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 929)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: Dr. James R. Lupski  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
CDNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM13596 row: f column: 12  
High quality sequence stop: 592.  
FEATURES  
Location/Qualifiers  
1. 929  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6193163"  
/sex="male"  
/tissue\_type="sympathetic trunk"  
/dev\_stage="adult, 16 yr"  
/lab\_host="DH10B"  
/clone\_lib="Lupski\_sympathetic trunk"  
/note="Vector: pCMV-SPORT6 (Life Technologies); Site\_1:  
NotI; Site\_2: SalI; cDNA made by oligo-dT priming.  
Directionally cloned using the following adaptors:  
5'-TCGACCCACGCGTCCG-3' and  
5'-GACTAGTCTAGATCGGCGGCCCT(15)-3'. Size selected >  
1 kb for average insert length 1.9 kb. This is a primary  
library, non-amplified. Library constructed by Life  
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor  
College of Medicine); available through Life  
Technologies."  
ORIGIN  
Query Match 20.2%; Score 746.2; DB 13; Length 929;  
Best Local Similarity 90.4%; Pred. No. 2.4e-113;  
Matches 831; Conservative 0; Mismatches 83; Indels 5; Gaps 3;  
QY 1412 AGTCATTGTGACGCCACCGCGCGGAGAGCCAGGTGCGGACGTGGCGTGGATCGGCG 1471  
Db 1 AGTCATTGTGACGCCACCGCGCGGAGAGCCAGGTGCGGACGTGGCGTGGATCGGCG 60

```
QY 1472 ATGGCGTATGGGTGTCCATCCGCTGGACTCCACCTTGAGGCTCTACCATGCACACACGC 1531
Db 61 ATGGCGTATGGGTGTCCATCCGCTGGACTCCACCTTGAGGCTCTACCATGCACACACGC 120

QY 1532 ACCAGCATCTACAGGACGTGGACATTGAGCCCTACGTACGCAAGATGCTAGGCACTGGCA 1591
Db 121 ACCAGCATCTACAGGACGTGGACATTGAGCCCTACGTACGCAAGATGCTAGGCACTGGCA 180

QY 1592 AGCTGGGTTTCTCCTTCGTACGCATCACGGGCTGCTTGTGCGGGCAGCGGCTCTGGG 1651
Db 181 AGCTGGGTTTCTCCTTCGTACGCATCACGGGCTGCTTGTGCGGGCAGCGGCTCTGGG 240

QY 1652 TGGGCACCGGCAACGGAGTGTGTCATCTCCATCCCTTGACAGAGACTGTGTCCTGCACC 1711
Db 241 TGGGCACCGGCAACGGAGTGTGTCATCTCCATCCCTTGACAGAGACTGTGTCCTGCACC 300

QY 1712 GAGGCCAGCTCCTGGGGCTCCGAGCCCAATAAGACATCCCCACCTCTGGGAGGGCGCCC 1771
Db 301 GAGGCCAGCTCCTGGGGCTCCGAGCCCAATAAGACATCCCCACCTCTGGGAGGGCGCCC 360

QY 1772 GTCCCCGGGGCATCATCCACGTGTATGGCGATGACAGCAGTGACAGGGCGGCAGCAGCT 1831
Db 361 GTCCCCGGGGCATCATCCACGTGTATGGCGATGACAGCAGTGACAGGGCGGCAGCAGCT 420

QY 1832 TCATCCCCCTACTGCTCCATGGCCAGGCCAGCTATGCTTCCATGGGCACCGGATGCCG 1891
Db 421 TCATCCCCCTACTGCTCCATGGCCAGGCCAGCTATGCTTCCATGGGCACCGGATGCCG 480

QY 1892 TGAAGTTCTTTGTCGGTGCCAGGAAACGTGTGGCCACCCCTGAAATGGCAGTGTGCTGG 1951
Db 481 TGAAGTTCTTTGTCGGTGCCAGGAAACGTGTGGCCACCCCTGAAATGGCAGTGTGCTGG 540

QY 1952 ACAGCCAGCCGAGGCGCCCTGGGCCAGCTGCCCTCGCTCGAGGTCGAGGGCCAGAAGC 2011
Db 541 ACAGCCAGCCGAGGCGCCCTGGGCCAGCTGCCCTCGCTCGAGGTCGAGGGCCAGAAGC 600

QY 2012 TGGCGAAGCTGCTGTGCTGAGCGGGGGAGGGCTACATCGACTTCCGCAATTGGAGACG 2071
Db 601 TGGCGAAGCTGCTGTGCTGAGCGGGGGAGGGCTACATCGACTTCCGCAATTGGAGACG 660

QY 2072 GAGAGGACGACGAGACGGAGGAGGGCGCAGGGGACATGAGCCAGGTGAAGCCCGTGTGT 2131
Db 661 GAGAGGACGACGAGACGGAGGAGGGCGCAGGGGACCTGAACCAGGTGAACCCCGTGTGT 720

QY 2132 CCAAGGCAGAGCGCAGTCACATCAT - CGTGTGGCAGGTGTCTTACACCCCGAGTGAAG 2189
Db 721 CCAAGGCAAAGCGCAAGTCCATCATCCCGGGGCGAGGGGCTTAACCCCGGAAGGAAC 780

QY 2190 CTGCTGCCCTGCCCTGG - CCGACCTGTACATAGG - ACCCCGACACACCTGACCCCGCC 2246
Db 781 CTCCTGCCCTTGCTGGGCCCGGACTCGAGCGTAGGAACCCCGGAACACCCCTGACCCCGC 840

QY 2247 CGGCCCGGGGTAGCCAGCCAGGCGCGCGCCCTCTTCTAACCTCTCAACCTGCAGC 2306
Db 841 CCAGGCCCCCGAGGTAACCCAGGCGGGGGTCCCCACGCCCCCTTCTTAGACGTCACC 900

QY 2307 TTTTCACTGAGTCTGGCCC 2325
Db 901 ACCCGGCTGTTTTCACCC 919
```

```
RESULT 8
BX434805/c
LOCUS BX434805 901 bp mRNA linear EST 15-MAY-2003
DEFINITION BX434805 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
CS0DF011YK03 3-PRIME, mRNA sequence.
ACCESSION BX434805
VERSION BX434805.1 GI:30779330
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 901)  
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 8197.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0BAK032BE07NM1&cluster=8197.f. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ InVitroGen Corporation 1600  
Paraday Avenue Genoscope sequence ID : CS0BAK032BE07NM1.

FEATURES  
source

Location/Qualifiers  
1. .901  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DF011YK03"  
/tissue\_type="FETAL BRAIN"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL BRAIN"  
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."

ORIGIN

Query Match 19.5%; Score 721.4; DB 13; Length 901;  
Best Local Similarity 90.5%; Pred. No. 3.1e-109;  
Matches 815; Conservative 0; Mismatches 81; Indels 5; Gaps 4;  
  
QY 2781 CCCTGCCCCAGGAGCTGGCATAAAGACAGAGCCCGCTCCTTGGGGCAGCTCTGA 2840  
Db 896 CCGGGCCCCAAAAGTTGGCAAGAAAGGCCGAGCCCCCGTTCTTGGGGGAAGATGTTA 837  
  
QY 2841 GAAACAGAGACTGTACCCCATCTGTCCTGATGAGCAGGCTCTTCCAGACCCCGTCTGA 2900  
Db 836 AGAAAAAATATGTTTCCCTTCAGGCCCCAGCAGCAGGTTACTGGCAAGCCCATGTTGA 777  
  
QY 2901 CCCTGTCCCCCAGGCTCTGCTGGGCAGAACTACCTTGGAGGAGTGGGCCCTGGA 2960  
Db 776 -CCGTGTACCCACAGGGTCTGCCGGGGCAGAAACTCA - CTTGGAGGAATGGGCCCGGAA 719  
  
QY 2961 GTCTGTCCCTCCAGAAAGCCCCCAGGGTGGATTTCTCAGGCTGCCAGGGCAGGCCAG 3020  
Db 718 TTAGTGTCTTCCCAAAAGCCCCAGGGT - GGATTTCTCAGGCTGCCAGGGCAGGCCAG 660  
  
QY 3021 GCCTCAGGAAGAGGGGAGGCCCTTGCCCTCTCCGGGATCAGTCTCCTAGGACACAGGCTCA 3080  
Db 659 GCCTCAGGAAGAGGGGAGGGCCCCCTGGCCCTCTCCGGGATAGAGCCTAGGACACAGGCTCA 600  
  
QY 3081 GCCTCAGGTTGATGGGGGATGATGTCTCCCGGGGCTGCCTCCTGACCGGGCTCCACG 3140  
Db 599 AGCTCAGGTTGATGGGGGATGATGTCTCCGGGGGCTGCCTCCTGACCGGGCTCCACG 540  
  
QY 3141 GAGCCAGCTCCACACACGCTACTAAGTGCCTAGGGTTGCCCGCTGTGGCCTGCTCCCA 3200  
Db 539 GAGCCAGCTCCACACACGCTACTAAGTGCCTAGGGTTGCCCGCTGTGGCCTGCTCCCA 480  
  
QY 3201 GGGAGCAACAGAGAGGCCCAACAGCAGAGGCCCGTGGGGCTGAGGATGGAGCCGCCCA 3260  
Db 479 GGGAGCAACAGAGAGAGGCCCAACAGCAGAGGCCCGTGGGGCTGAGGATGGAGCCGCCCA 420  
  
QY 3261 GCCGACTCCAAAGCCCGCAGAGGGGACAGCCACCCCTGACTGCTCTCCCTGCCAGCTGG 3320  
Db 419 GCCGACTCCAAAGCCCGCAGAGGGGACAGCCACCCCTGACTGCTCTCCCTGCCAGCTGG 360  
  
QY 3321 GCCTCTGTGGCCTATTCTTCTACCTTCCAGGCCCACTGCACTCCTGTCTGGGAGGCCCTTAT 3380

|||||  
Db 359 GCCTATCTGGCCTATCTCTACCTTCCAGGCCCACTGCACTCCTGTCTGGGAGGCCCTTAT 300  
3381 GAGGGCAGCCCGCCCGCCACCCACCCCAACACAGAGAAACACAGATCTTGGGAGCTG 3440  
Db 299 GAGGGCAGCCCGCCCGCCACCCACCCCAACACAGAGAAACACAGATCTTGGGAGCTG 240  
QY 3441 CCCCACAAGCCCGCTGGCCACCCAGGGCTGCAGCGCTGCGCTGCCGGCTTCTCCCCAC 3500  
Db 239 CCCCACAAGCCCGCTGGCCACCCAGGGCTGCAGCGCTGCGCTGCCGGCTTCTCCCCAC 180  
QY 3501 CACCCTGCCACCTCCACTGTGATGTATGTCCGCTCCCTCGTCTGTTCCCGCAGGATCTCG 3560  
Db 179 CACCCTGCCACCTCCACTGTGATGTATGTCCGCTCCCTCGTCTGTTCCCGCAGGATCTCG 120  
QY 3561 AAGTGACTCCGGGCTGAGCAGTGCGGCGGCTGGGGAGGGGTGACGATTCTCCTCAGGCT 3620  
Db 119 AAGTGACTCCGGGCTGAGCAGTGCGGCGGCTGGGGAGGGGTGACGATTCTCCTCAGGCT 60  
QY 3621 TTGGCCCTGCAAGCAACCCACATATCTGCTGTATGTATTAATAATGTTTAACGTCGTA 3680  
Db 59 TTGGCCCTGCAAGCAACCCACATATCTGCTGTATA--GAAATAAGGTCTTAACGTCGAA 2  
QY 3681 A 3681  
1 A 1

RESULT 9  
BX451957  
LOCUS  
DEFINITION BX451957 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone  
CS0DF011YK03 5-PRIME, mRNA sequence.  
ACCESSION BX451957  
VERSION BX451957.1 GI:31022315  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 981)  
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 8197.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0BAG016ZB07\_CS01490\_2&cluster=8197.f.  
Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0BAG016ZB07\_CS01490\_2.  
Location/Qualifiers  
1. .981  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DF011YK03"  
/tissue\_type="FETAL BRAIN"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL BRAIN"  
/note="Organ: brain; Vector: pCMVSPORT\_6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."

ORIGIN

Query Match

19.4%; Score 717; DB 13; Length 981;

Best Local Similarity 89.0%; Pred. No. 1.7e-108;  
Matches 820; Conservative 0; Mismatches 95; Indels 6; Gaps 4;  
QY 2646 GGGAGCTGGGCGAGGCACTAGCTTTGCCCCAGGAGGTGGGCTCAGGCTGCCAGGTG 2705  
Db 42 GGAAGCTGGGTCGGCACTAGCTTTGCCCCAGGAGGTGGGCTCAGGCTGCCAGGTG 101  
QY 2706 CCTGACCCCGAGCGGCGCTTCTTGGGGCTTCCCCGTGTCAGGCTCTATCCTGCTGT 2765  
Db 102 CTTGACCCCGAGCGGCGCTTCTTGGGGCTTCCCCGTGTCAGGCTCTATCCTGCTGT 161  
QY 2766 CCCCACCCAGCTGTCCCTGCCCCAGGAGCTTGGCATAAAAACAGGAGGCGGCTCCCT 2825  
Db 162 CCCCACCCAGCTGTCCCTGCCCCAGGAGCTTGGCATAAAAACAGGAGGCGGCTCCCT 221  
QY 2826 GGGGAGCTGCTTGAGAACAGAGACTGCTACCCCATCTGCCATGCAGGAGGCTCTTG 2885  
Db 222 GGGGAGCTGCTTGAGAACAGAGACTGCTACCCCATCTGCCATGCAGGAGGCTCTTG 281  
QY 2886 CCAGCCCCGTTTGACCCGTTGTCCTCCAGGCTTCCCTGGGAGAGACTCACCTTGA 2945  
Db 282 CCAGCCCCGTTTGACCCGTTGTCCTCCAGGCTTCCCTGGGAGAGACTCACCTTGA 341  
QY 2946 GGAGTGGGCGCTGGAGTCTGTCTCCCAAGAGGCGGCTTCCCTGCGGATCAGTCC 3065  
Db 342 GGAGTGGGCGCTGGAGTCTGTCTCCCAAGAGGCGGCTTCCCTGCGGATCAGTCC 401  
QY 3006 CCAGGGCAGGCGCTCAGGAAGAGGAGGCGGCTTCCCTGCGGATCAGTCC 3065  
Db 402 CCAGGGCAGGCGCTCAGGAAGAGGAGGCGGCTTCCCTGCGGATCAGTCC 461  
QY 3066 TAGGACAGGCTCAGCTCAGGTTGATGGGGGATGATGTCTCCCGGGCTGCTCCT 3125  
Db 462 TAGGACAGGCTCAGCTCAGGTTGATGGGGGATGATGTCTCCCGGGCTGCTCCT 521  
QY 3126 GCACGGGCTCCACGGAGCCAGCTCCAGACACAGCTACTAAGTGCCTAGGTTG-CCCG 3184  
Db 522 GC-CGGGGCTCCACGGAGCCAGCTTCCAGACACGCTACTAAGTGCCTAGGTTGCCCT 580  
QY 3185 CTGTGGCTGCTCCAGGGAGCAACAGAGAGGCCACCAAGCAGAGGCGGCTGGGCTGAG 3244  
Db 581 CTGTGGCTGCTCCAGGGAGCAACAGAGAGGCCACCAAGCAGAGGCGGCTGGGCTGAA 640  
QY 3245 GATGAGCGGCGGCGGAGCTCCAGCGGCTCCAGAGCGGAGGCGGCTGCGGCTGCT 3304  
Db 641 GATGAGCGGCGGCGGAGCTCCAGCGGCTCCAGAGCGGAGGCGGCTGCGGCTGCT 700  
QY 3305 CTCCCTGCGGCTGGG--CCTCTCTGCGCTTATTCCTACCTTCCAGGCGGCTGCTCC 3362  
Db 701 TTTCTTGGCAGGTTGGGCGCTTTGTGGCCATTTCCTAATTTTCAGGCGGCTGCT 760  
QY 3363 TGTCTGGAGGCGCTTATGAGGCGGAGCGGCGGCTCCAGCGGCGGCTGCGGCTGCGG 3422  
Db 761 TGTATGGGAAGCGCTTATGAAGGGAGCGGCGGCTCCGAACCCACCCCAACAGAGAG 820  
QY 3423 ACAGATCTTGGGAGCTGCGGCGGCTCCAGCGGCTGCGGCTGCGGCTGCGGCTGCGG 3482  
Db 821 ACCAATATTGGGAGTGTCCCAACAGCGGCTTGCAGGCGGCTGCGGCTGCGGCTGAG 880  
QY 3483 CTGCGGCTTCTCCCGACCCCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 3542  
Db 881 TTGACGGTTTTTGCCCA--ACCTTGGCCACTTACTTGGGAGATATGGCGGTTCCCTGCGG 938  
QY 3543 TGTTCCTCCCGAGGATCTCGAAG 3563  
Db 939 TGTTCCTCCCGAAGGATG 959

RESULT 10  
BI767677

LOCUS

DEFINITION

BI767677 748 bp mRNA linear EST 25-SEP-2001  
603060624F1 NIH\_MGC\_122 Homo sapiens cDNA clone IMAGE:5209972 5',  
mRNA sequence.

ACCESSION BI767677  
VERSION BI767677.1 GI:15759255  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 748)  
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLAM11526 row: p column: 05  
High quality sequence stop: 746.  
Location/Qualifiers  
1..748  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5209972"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_122"  
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleens. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH\_MGC Library."  
ORIGIN  
Query Match 19.3%; Score 713.4; DB 12; Length 748;  
Best local Similarity 99.5%; Pred. No. 6.3e-108;  
Matches 747; Conservative 0; Mismatches 1; Indels 3; Gaps 3;  
1125 CTCGGTGTGGCCAACTGGAAGAGCGCCTGCATCCATCAAGCTGAAGGATTCTGTGCT 1184  
1 CTCGGTGTGGCCAACTGGAAGAGCGCCTGCATCCATCAAGCTGAAGGATTCTGTGCT 60  
1185 GAGCCTGGTGCATGTCAAAGGCCGTGTGCTGGTGGCTCTGGCGGACGGACCCCTGGCCAT 1244  
61 GAGCCTGGTGCATGTCAAAGGCCGTGTGCTGGTGGCTCTGGCGGACGGACCCCTGGCCAT 120  
1245 CTTCCACCGTGGTGAAGATGGCCAGTGGGATCTGAGCAACTATCACCTAATGACCTGGG 1304  
121 CTTCCACCGTGGTGAAGATGGCCAGTGGGATCTGAGCAACTATCACCTAATGACCTGGG 180  
1305 CCACCCGACCACTCCATCCGCTGCATGGCTGTTGTGTACGACCGCGTGTGGTGTGGCTA 1364  
181 CCACCCGACCACTCCATCCGCTGCATGGCTGTTGTGTACGACCGCGTGTGGTGTGGCTA 240  
1365 CAAGAACAGGTGCACGTCACTCCAGCCCAAGACCATGCGATAGAGAGTCAATTTGACGC 1424  
241 CAAGAACAGGTGCACGTCACTCCAGCCCAAGACCATGCGATAGAGAGTCAATTTGACGC 300  
1425 CCACCCGCGCGGGAGAGCCAGGTGCGGCAGCTGGCGTGGATCGGCGATGGCGTATGGGT 1484  
301 CCACCCGCGCGGGAGAGCCAGGTGCGGCAGCTGGCGTGGATCGGCGATGGCGTATGGGT 360  
1485 GTCCATCCGCTGGACTCCACCCCTGAGGCTCTACCATGCACACACGACCGACCATCTACA 1544  
361 GTCCATCCGCTGGACTCCACCCCTGAGGCTCTACCATGCACACACGACCGACCATCTACA 420

QY 1545 GGACGTGGACATTGAGCCCTACGTAGCAAGATGCTAGGACCTGGCAAGTGGGTTTCTC 1604  
Db 421 GGACGTGGACATTGAGCCCTACGTAGCAAGATGCTAGGACCTGGCAAGTGGGTTTCTC 480  
QY 1605 CTTCTGTACGCATCACGGCCCTGTTGTGCGGGGAGCCGGCTCTGGGTGGGACCGGCAA 1664  
Db 481 CTTCTGTACGCATCACGGCCCTGTTGTGCGGGGAGCCGGCTCTGGGTGGGACCGGCAA 540  
QY 1665 CGGAGTGGTCACTCTCCATCCCCCTGACAGAGACTGTGGTCTGCACCGAGGCGAGCTCCT 1724  
Db 541 CGGAGTGGTCACTCTCCATCCCCCTGACAGAGACTGTGGTCTGCACCGAGGCGAGCTCCT 600  
QY 1725 GGGGCTCCGAGCCAATAAGACATCCCCACCTCTGGGGAGGGCGCCCGTCCCGGGGGCAT 1784  
Db 601 GGGGCTCCGAGCCAATAAGACATCCCCCA-CTCTGGGGAGGGCGCCCGTCCCGGGGGCAT 659  
QY 1785 CATCCACGTGTATGGCGATGACAGCAGTGAAGGGCGGCGGAGCAGCTTATCCCTACTG 1844  
Db 660 CATCCACGTGTATGGCGATGACAGCAGTGAAGGGCGG-CAGCAGCTTATCCCTACTG 718  
QY 1845 CTCCATGGCCAGGCCCGCAGCTATGCTTCCAT 1875  
Db 719 CTCCATGGCCAAAGG-CCAGCTATGCTTCCAT 748  
RESULT 11  
CD365632/c  
LOCUS  
DEFINITION UI-H-FT2-bjk-a-24-0-UI.s1 NCI CGAP FT2 Homo sapiens cDNA clone  
UI-H-FT2-bjk-a-24-0-UI 3', mRNA sequence.  
ACCESSION CD365632  
VERSION CD365632.1 GI:31149722  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 690)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Dr. Gary W. Hunninghake, U of I  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at <http://genome.uiowa.edu/distribution/cgap.html>  
Seq primer: M13 FORWARD  
POLYA=Yes.  
Location/Qualifiers  
1..690  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-H-FT2-bjk-a-24-0-UI"  
/tissue\_type="Aveolar Macrophage"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI-CGAP FT2"  
/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I;  
NCI CGAP FT2 is a subtracted cDNA library constructed from a pool of 81 RNA samples from Alveolar Macrophages challenged with different treatments. The library was subtracted according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. The tissue was provided by Dr. Gary W. Hunninghake of the University of Iowa.  
TAG TISSUE=Human Lung Aveolar Macrophage  
TAG\_LIB=UI-H-FT2



```
QY      2425 GCAGCTCTGGCCAGCTTCAGGCCAGAGTCTCAAGTCCAGGGCACCTTGGGCCCCAGCG 2484
      |||||||
Db      601 GCAGCTCTGGCCAGCTTCAGGCCAGAGTCTCAAGTCCAGGGCACCTTGGGGCCCCAGC 660
      |||||||
QY      2485 CAGGCAGAAAT-CCGAGGTGCTCTGGCTTACCC--TGGGCTCCTACTCCCCAGCACCC 2541
      |||||||
Db      661 CAGGCAGAAATCCCAAGGTGCTCTGGCCCTCCCTGGGGCTCCTACTCCCATCACCC 720
      |||||||
QY      2542 C 2542
      |
Db      721 C 721

RESULT 13
BI913764
LOCUS      BI913764
DEFINITION      603180726F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5244630 5',
      mRNA sequence.
ACCESSION      BI913764
VERSION      BI913764.1 GI:16178127
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (bases 1 to 721)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
      Email: cgapbs-r@mail.nih.gov
      Tissue Procurement: Life Technologies, Inc.
      cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
      DNA Sequencing by: Incyte Genomics, Inc.
      Clone distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LLNL at:
      http://image.llnl.gov
      Plate: LLAM11617 row: d column: 07
      High quality sequence stop: 716.
FEATURES
      source
      1. .721
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:5244630"
      /lab_host="DH10B"
      /clone_lib="NIH_MGC_121"
      /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
      Site 2: EcoRV (destroyed); RNA source anonymous pool of 3
      fetal brains, female age 20 weeks, female age 24 weeks,
      and male age 26 weeks. Library is oligo-dT primed and
      directionally cloned (EcoRV site is destroyed upon
      cloning). Average insert size 1.7 kb, insert size range
      0.7-3.5 kb. Library is normalized and enriched for
      full-length clones and was constructed by C. Gruber
      (Invitrogen). Research Genetics tracking code 017. Note:
      this is a NIH_MGC Library."
ORIGIN
Query Match      18.0%; Score 665; DB 12; Length 721;
Best Local Similarity      98.8%; Pred. No. 6.4e-100;
Matches 712; Conservative      0; Mismatches      5; Indels      4; Gaps      4;

QY      1787 TCCACGTGTATGGCGATGACAGCAGTGCAGGGCGGCCAGCAGCTTCATCCCTACTGCT 1846
      |||||||
Db      1 TCCACGTGTATGGCGATGACAGCAGTGCAGGGCGGCCAGCAGCTTCATCCCTACTGCT 60
      |||||||
QY      1847 CCATGGCCAGGCCAGCTATGCTTCCATGGGCACCGCGATGCCGTGAAGTCTTTGTCT 1906
      |||||||
Db      61 CCATGGCCAGGCCAGCTATGCTTCCATGGGCACCGCGATGCCGTGAAGTCTTTGTCT 120
      |||||||
QY      1907 CGGTGCCAGGAACGTGCTGGGCCACCCTGAATGGCAGTGTGCTGGACAGCCAGCCGAGG 1966
      |||||||
```

```
Db      121 CGGTGCCAGGAACGTGCTGGGCCACCCTGAATGGCAGTGTGCTGAGCCAGCCGAGG 180
      |||||||
QY      1967 GCCCTGGCCAGCTGCCCTGCTCGGAGGTTCGAGGCGCAGAAAGCTCGGAAACGTGCTGG 2026
      |||||||
Db      181 GCCCTGGCCAGCTGCCCTGCTCGGAGGTTCGAGGCGCAGAAAGCTCGGAAACGTGCTGG 240
      |||||||
QY      2027 TGCTGAGCGCGGGGAGGGGTACATCGACTTCCGATTGGAGACGGAGAGCAGCAGAGA 2086
      |||||||
Db      241 TGCTGAGCGCGGGGAGGGGTACATCGACTTCCGATTGGAGACGGAGAGCAGCAGAGA 300
      |||||||
QY      2087 CGGAGGAGGGCGCAGGGGACATGAGCCAGGTGAAGCCCGTGTGTCCAAAGCAGAGCGCA 2146
      |||||||
Db      301 CGGAGGAGGGCGCAGGGGACATGAGCCAGGTGAAGCCCGTGTGTCCAAAGCAGAGCGCA 360
      |||||||
QY      2147 GTCACATCATCGTGTGGCAGGTGCTCTACACCCCGAGTGAAGCTGCTGCCCTGCTGGC 2206
      |||||||
Db      361 GTCACATCATCGTGTGGCAGGTGCTCTACACCCCGAGTGAAGCTGCTGCCCTGCTGGC 420
      |||||||
QY      2207 CCGACCTGTACATAGGACCCCGGAGCACCTGACCCCGTCCGGCCCGGGGTAGCCAGC 2266
      |||||||
Db      421 CCGACCTGTACATAGGACCCCGGAGCACCTGACCCCGTCCGGCCCGGGGTAGCCAGC 480
      |||||||
QY      2267 CAGGCGCGCGCCCTCTTCTAACTCTCAACCTGCAGCTTTCACCTGAGTCTGGCCCC 2326
      |||||||
Db      481 CAGGCGCGCGCTGCTCTTCTAACTCTCAACCTGCAGCTTTCACCTGAGTCTGGCCCC 540
      |||||||
QY      2327 TCCAGCGGGCAGGAGTGCAGGATCGGATCAGCTGGAGGAGGGGAGGGGTGCTT 2386
      |||||||
Db      541 TCCAGCGGGCAGGAGTGCAGGATCGGATCAGCTGGAGGAGGGGAGGGGTGCTT 600
      |||||||
QY      2387 CCACCCGAGGGG-AAGATGCTCTCGGGACAG-TTCCCGGGCAGCTCCTGGCC-AGCTTC 2443
      |||||||
Db      601 CCACCCGAGGGGAAAGATGCTCTCGGGACAGATTACCCGGGAGCTCCTGGCCGAGCTTC 660
      |||||||
QY      2444 CAGCCCGAGGTCTCAAGTCCAGGGCACCTTGGGGCCCGCAGCGCAGGAGGAGTCCGAGGTG 2502
      |||||||
Db      661 CAGCCCGAGGTCTCAAGTCCAGGGCACCTTGGGGCCCGCAGCGCAGGAGGAGTCCGAGGTG 720
      |||||||
QY      2503 G 2503
      |
Db      721 G 721

RESULT 14
BI861206
LOCUS      BI861206
DEFINITION      603389971F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5399224 5',
      mRNA sequence.
ACCESSION      BI861206
VERSION      BI861206.1 GI:16001953
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (bases 1 to 727)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
      Email: cgapbs-r@mail.nih.gov
      Tissue Procurement: DCTD/DTP
      cDNA Library Arrayed by: Life Technologies, Inc.
      DNA Sequencing by: Incyte Genomics, Inc.
      Clone distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LLNL at:
      http://image.llnl.gov
      Plate: LLAM12017 row: m column: 17
      High quality sequence stop: 708.
FEATURES
      source
      1. .727
```

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5399224"  
/tissue\_type="mammary adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC 87"  
/note="Organ: breast; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: Sall; Cloned unidirectionally; oligo-dr primed.  
Average insert size 1.383 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 18.0%; Score 664.2; DB 12; Length 727;  
Best Local Similarity 98.9%; Pred. No. 8.7e-100;  
Matches 721; Conservative 0; Mismatches 3; Indels 5; Gaps 5;  
  
QY 839 CFTCCAGTCCACAGAGGAGGCCACAGAGGCCACGAGGTGCCAGACCCCTGGGCCAGCG 898  
Db |||||  
1 CGTCCAGTCCACAGAGGAGGCCACAGAGGCCACGAGGTGCCAGACCCCTGGGCCAGCG 60  
  
QY 899 AGCCA-GAGACAGCCACATTGCGGCCCGGGCTCTCACAGAGCAGTCTTCACTGACCCA 957  
Db |||||  
61 AGCCATGAGACAGCCACATTGCGGCCCGGGCTCTCACAGAGCAGTCTTCACTGACCCA 120  
  
QY 958 GCCCCGACCCCTCTCTGCCCCCAGCCTGCGAGCGAGAACGGGCCAGAGCTGACAGC 1017  
Db |||||  
121 GCCCCGACCCCTCTCTGCCCCCAGCCTGCGAGCGAGAACGGGCCAGAGCTGACAGC 180  
  
QY 1018 AGCAGCACACGCCCA-GAGCCAGAGCCCGGGGACCCCGAGGAGCAGGAGCAGTGC 1076  
Db |||||  
181 AGCAGCACACGTCATGAGCCAGAGCCCGGGGACCCCGAGGAGCAGGAGCAGTGC 240  
  
QY 1077 TGCACCCACCATGTGGCTGGAGCCCGAGAGCGGCTGCTATGTGCACTCGGCTGTGC 1136  
Db |||||  
241 TGCACCCACCATGTGGCTGGAGCCCGAGAGCGGCTGCTATGTGCACTCGGCTGTGC 300  
  
QY 1137 CAACTGGAAGAAGTGCCTGCACTCCATCAAGCTGAAGATCTGTGTGAGCTGTGCA 1196  
Db |||||  
301 CAACTGGAAGAAGTGCCTGCACTCCATCAAGCTGAAGATCTGTGTGAGCTGTGCA 360  
  
QY 1197 TGTCAAAGCCGCTGTGCTGGCTCTGGCGGACCGGACCTGGCCATCTTCCACCGTGG 1256  
Db |||||  
361 TGTCAAAGCCGCTGTGCTGGCTCTGGCGGACCGGACCTGGCCATCTTCCACCGTGG 420  
  
QY 1257 TGAAGATGCCAGTGGGATCTGAGCAACTATCACTATGACCTGGGCCACCCGACCA 1316  
Db |||||  
421 TGAAGATGCCAGTGGGATCTGAGCAACTATCACTATGACCTGGGCCA-CCGACCA 479  
  
QY 1317 CTCCATCCGCTGCATGGTGTGTGTACGACCGGTGTGTGTGCTGAGCAAGCAAGGT 1376  
Db |||||  
480 CTCCATCCGCTGCATGGTGTGTGTACGACCGGTGTGTGTGCTGAGCAAGCAAGGT 539  
  
QY 1377 GCACGTATCCAGCCCAAGACCATGCAGATAGAGAGTCAATTTGACGCCACCCCGCGG 1436  
Db |||||  
540 GCACGTATCCAGCCCAAGACCATGCAGATAGAGAGTCAATTTGACGCCCA-CCGCGGCG 598  
  
QY 1437 GGAGAGCCAGGTGCGGAGCTGGCGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCT 1496  
Db |||||  
599 GGAGAGCCAGGTGCGGAGCTGGCGTGGATCGGCGATGGCGTATGGGTGTCCATCCGACT 658  
  
QY 1497 GGACTCCACCTGAGGCTCTACCATGCACACACGACAGCATCTACAGGA-CGTGGACA 1555  
Db |||||  
659 GGACTCCACCTGAGGCTCTACCATGCACACACGACAGCATCTACAGGACCGTGGACA 718  
  
QY 1556 TTGAGCCCT 1564  
Db |||||  
719 TAGAGCCCT 727

RESULT 15  
BU614626

LOCUS BU614626 862 bp mRNA linear EST 20-FEB-2003  
DEFINITION UI-M-EV0-cbh-m-09-0-UI.r1 NIH BMAP EV0 Mus musculus cDNA clone  
UI-M-EV0-cbh-m-09-0-UI 5', mRNA sequence.  
ACCESSION BU614626  
VERSION BU614626.1 GI:23280841  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 862)  
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Dr. James Lin, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, bento-soares@uiowa.edu  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

Seq primer: pYX-5.

Location/Qualifiers  
1. .862  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="UI-M-EV0-cbh-m-09-0-UI"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 15.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH BMAP EV0"  
/note="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured mRNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with an  
oligo-dr primer containing a Not I site. Double stranded  
cDNA was size selected according to mRNA size fraction,  
ligated with EcoR I adaptor, digested with Not I, and then  
cloned directionally into pYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA  
tail, is GTGGTGGAA. This library was created for the  
University of Iowa Mouse Brain Molecular Anatomy Project  
(BMAP). 'Gene Discovery in the Developing Mouse Nervous  
System', supported by National Institutes of Mental Health  
(NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN

Query Match 17.9%; Score 660; DB 13; Length 862;  
Best Local Similarity 85.5%; Pred. No. 4.5e-99;  
Matches 735; Conservative 0; Mismatches 125; Indels 0; Gaps 0;  
  
QY 709 GGTATCAACCTGGTGGGCTGTGCCACCCCGCTGCAAGCTGCCGCGAGCAACTGCTCTCC 768  
Db |||||  
2 GGCATCACCTGGTGGGCTGTGTACCCGCTGCAATGTTCCACGTAGCAACTGTTCTCTCA 61  
  
QY 769 CGAGGGACACCCCGTCTAGACAGGGGCGAGGGAGGTGGCCACCATCGCCACGGG 828  
Db |||||  
62 CGAGAGACACCCCGTCTAGACAGGGGCGAGGGAGGTGGCCACCATCGCCATGGG 121  
  
QY 829 AAGTCAACCCCGTCTCCAGTCCACAGAGGAGGCCACAGAGCGGAGGTGCCAGCCCT 888  
Db |||||  
122 AAGTCAACCCCGTCTCCAGTCCACAGAGGAGGCCACAGAGCGGAGGTGCCAGCCCT 181  
  
QY 889 GGGCCAGCGAGCGAGAGCAGCCACATTGCGGGCCCGGCTCTTCACAGAGCAGTCTTC 948  
Db |||||  
182 GGTCCAGCGAGTCAAGAGCAGCAGTCCGCGCCCGGCTCTTCACAGAGCATGTCTTT 241

|    |      |   |      |
|----|------|---|------|
| QY | 949  | ACTGACCCAGCCCGACCCCGTCCTCTGGCCCCCAGCCTGGCAGCGAGAACGGGCCAGAG   | 1008 |
| Db | 242  | ACTGACCCAGCACCCACCCCATCTCCAGCACCCAGCCTGCCAGTGAGATGGGTGAGAG    | 301  |
| QY | 1009 | CCTGACAGCAGCAGCACACGGCCAGAGCCAGAGCCCGGGGACCCACGGGAGCAGGC      | 1068 |
| Db | 302  | TCCAATGGCACCATTGTACAGCCTCAGGTGAGGCCAGTGGGGAACCTCTCAACAACAACC  | 361  |
| QY | 1069 | AGCAGTGTCTGACCCACCATGTGGCTGGGAGCCAGAACGGCTGGCTCTATGTGCACTCG   | 1128 |
| Db | 362  | AGTAGCGCTGACCCACATATGTGGCTAGGAGCCAGATGGCTGGCTCTATGTGCATTCA    | 421  |
| QY | 1129 | GCTGTGGCCAACTGGAAGAGTGCCTGCACCTCAAGCTGAAGGATTCTGTGCTGAGC      | 1188 |
| Db | 422  | GCGGTAGCCAACTGGAAGAGTGTCTGCACCTCAAGCTAAAAAGACTCTGTGCTGAGC     | 481  |
| QY | 1189 | CTGGTGCATGTCAAAGGCCGTGTGCTGGTGGCTCTGGCGGAGGGACCCCTGGCCATCTTC  | 1248 |
| Db | 482  | CTGGTGCATGTCAAAGGCCGTGTGCTGGTGGCTCTGGAGATGGGACCCCTGGCTATCTTC  | 541  |
| QY | 1249 | CACCGTGGTGAAGATGGCCAGTGGGATCTGAGCAACTATCACCTAATGGACCTGGGCCAC  | 1308 |
| Db | 542  | CATCGTGGAGAGGATGGCCAGTGGGACCTGAGCAACTACCACCTAATGGACCTGGGCCAC  | 601  |
| QY | 1309 | CCGACCACTCCATCCGCTGCATGGCTGTTGTGTACGACCGCGTGTGGTGTGGCTACAAG   | 1368 |
| Db | 602  | CCACACCACTCCATCCGCTGCATGGCTGTTGTGAATGACCGAGTTGGTGTGGCTACAAG   | 661  |
| QY | 1369 | AACAAGGTGCACCTCATCCAGCCCAAGACCATGCAGATAGAGAAGTCATTTGACGCCCCAC | 1428 |
| Db | 662  | AACAAGGTGCATGTTATCCAGCCCAAGACAATGCAGATTGAGAAATCATTTGATGCCCCAC | 721  |
| QY | 1429 | CCGCGCGGGAGAGCCAGGTCCGGCAGCTGGCGTGGATCGGCGATGGCGTATGGGTGTCC   | 1488 |
| Db | 722  | CCAAGCGGGAAGCCAGGTACGTCAGCTGGCCTGGATCGGTGATGGAGTGTGGGTCTCT    | 781  |
| QY | 1489 | ATCGCCTGGACTCCACCCCTGAGGCTCTACCATGACACACACGCCAGCATCTACAGGAC   | 1548 |
| Db | 782  | ATTGCTTGGATTCTACCCCTTCGSGCTCTACCATGCTCACACCCACAGCACCTGCAGGAT  | 841  |
| QY | 1549 | GTGGACATTGAGCCCTACGT  | 1568 |
| Db | 842  | GTGGACATTGAGCCCTATGT  | 861  |

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2004, 00:04:58 ; Search time 13735 Seconds  
(without alignments)  
11657.020 Million cell updates/sec

Title: US-10-019-495-26  
Perfect score: 3694  
Sequence: 1 ctggcagggttcctagtgagc.....gtcgtaaaaaaaaaaaaaa 3694

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg\_hum.\*

31: em.htg\_inv.\*

32: em.htg\_other.\*

33: em.htg\_mus.\*

34: em.htg\_pln.\*

35: em.htg\_rod.\*

36: em.htg\_mam.\*

37: em.htg\_vrt.\*

38: em.sy.\*

39: em.htgo\_hum.\*

40: em.htgo\_mus.\*

41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match % | Length | DB | ID        | Description         |
|------------|--------|---------------|--------|----|-----------|---------------------|
| 1          | 3694   | 100.0         | 3694   | 6  | BD269958  | BD269958 Human mem  |
| 2          | 3694   | 100.0         | 3694   | 6  | AX041037  | AX041037 Sequence   |
| 3          | 3484.6 | 94.3          | 4701   | 6  | AX440473  | AX440473 Sequence   |
| 4          | 3482   | 94.3          | 5621   | 9  | AB028989  | AB028989 Homo sapi  |
| 5          | 1816.2 | 49.2          | 2338   | 9  | AB071076  | AB071076 Macaca fa  |
| 6          | 1807   | 48.9          | 5440   | 10 | BC060603  | BC060603 Mus muscu  |
| 7          | 1797.4 | 48.7          | 5443   | 10 | AF178636  | AF178636 Mus muscu  |
| 8          | 1780.6 | 48.2          | 5411   | 10 | AB093282  | AB093282 Mus muscu  |
| 9          | 1770.4 | 47.9          | 5532   | 10 | AF178637  | AF178637 Mus muscu  |
| 10         | 1751.4 | 47.4          | 20612  | 9  | HS371H6   | AL031718 Human DNA  |
| 11         | 1751.4 | 47.4          | 270150 | 9  | AE006639  | AE006639 Homo sapi  |
| 12         | 1746.6 | 47.3          | 69437  | 9  | AC012180  | AC012180 Homo sapi  |
| 13         | 1542.6 | 41.8          | 4545   | 10 | AF262046  | AF262046 Mus muscu  |
| 14         | 1531.6 | 41.5          | 4173   | 10 | AB005662  | AB005662 Mus muscu  |
| 15         | 1497.6 | 40.5          | 3945   | 10 | AB043123  | AB043123 Mus muscu  |
| 16         | 1497.6 | 40.5          | 4011   | 10 | AB043125  | AB043125 Mus muscu  |
| 17         | 1497.6 | 40.5          | 4014   | 10 | AB043124  | AB043124 Mus muscu  |
| 18         | 520    | 14.1          | 1728   | 10 | BC004003  | BC004003 Mus muscu  |
| 19         | 505.2  | 13.7          | 5311   | 10 | BC060506  | BC060506 Mus muscu  |
| 20         | 463.2  | 12.5          | 2316   | 9  | AF542172  | AF542172 Homo sapi  |
| 21         | 442.2  | 12.0          | 4667   | 10 | AF327451  | AF327451 Mus muscu  |
| 22         | 440.6  | 11.9          | 4221   | 10 | BC060100  | BC060100 Mus muscu  |
| 23         | 439.4  | 11.9          | 182073 | 2  | AC1114907 | AC1114907 Mus muscu |
| 24         | 439.4  | 11.9          | 212298 | 2  | AC121120  | AC121120 Mus muscu  |
| 25         | 436.2  | 11.8          | 175759 | 10 | AF220294  | AF220294 Mus muscu  |
| 26         | 413.4  | 11.2          | 3825   | 9  | AB011088  | AB011088 Homo sapi  |
| 27         | 413.4  | 11.2          | 4675   | 9  | AF327452  | AF327452 Homo sapi  |
| 28         | 389    | 10.5          | 199278 | 2  | AC130925  | AC130925 Rattus no  |
| 29         | 389    | 10.5          | 239810 | 2  | AC111901  | AC111901 Rattus no  |
| 30         | 370.2  | 10.0          | 3113   | 6  | AR083288  | AR083288 Sequence   |
| 31         | 370.2  | 10.0          | 3113   | 6  | AX099558  | AX099558 Sequence   |
| 32         | 370.2  | 10.0          | 3113   | 6  | BD056704  | BD056704 Secreted   |
| 33         | 360.6  | 9.8           | 3791   | 6  | AX281594  | AX281594 Sequence   |
| 34         | 352.4  | 9.5           | 43288  | 9  | HS361A3   | AL031717 Human DNA  |
| 35         | 294.4  | 8.0           | 4239   | 3  | AF262045  | AF262045 Drosophil  |
| 36         | 294.4  | 8.0           | 4455   | 3  | BT001718  | BT001718 Drosophil  |
| 37         | 279    | 7.6           | 1966   | 6  | AX880664  | AX880664 Sequence   |
| 38         | 279    | 7.6           | 1966   | 6  | BD158516  | BD158516 Primer fo  |
| 39         | 279    | 7.6           | 1966   | 9  | AK023512  | AK023512 Homo sapi  |
| 40         | 252.8  | 6.8           | 1037   | 3  | AY060748  | AY060748 Drosophil  |
| 41         | 244    | 6.6           | 633    | 6  | AX869408  | AX869408 Sequence   |
| 42         | 244    | 6.6           | 633    | 6  | BD149470  | BD149470 Primer fo  |
| 43         | 237    | 6.4           | 237    | 6  | AX884437  | AX884437 Sequence   |
| 44         | 237    | 6.4           | 237    | 6  | BD024047  | BD024047 Sequence   |
| 45         | 220.4  | 6.0           | 10938  | 2  | AC014510  | AC014510 Drosophil  |

ALIGNMENTS

RESULT 1  
BD269958  
LOCUS BD269958 3694 bp DNA linear PAT 17-JUL-2003  
DEFINITION Human membrane-associated proteins.  
ACCESSION BD269958  
VERSION BD269958.1 GI:33079726  
KEYWORDS JP 2002542782-A/9.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 3694)  
AUTHORS Hillman,J.L., Bandman,O., Tang,T.Y., Lal,P., Yue,H., Reddy,R.,  
Azimzai,Y. and Baughn,M.R.  
TITLE Human membrane-associated proteins

|                            |  |     |      |   |      |
|----------------------------|--|-----|------|---|------|
| JOURNAL                    | Patent: JP 2002542782-A 9 17-DEC-2002;                                   | Db  | 421  | AAGAGCGCCACACGTCCTCCGAGAAAGAAAGGCCAAGGAGCTCCCTGAAATGGACGCC      | 480  |
| COMMENT                    | INCYTE GENOMICS INC  | QY  | 481  | ACCTCCAGCCGGGTGTGGATCCTGTACCAAGCACCCCTGACCAACAGCAAGGTGGTGATCATC | 540  |
|                            | OS Homo sapiens (human)  | Db  | 481  | ACCTCCAGCCGGGTGTGGATCCTGTACCAAGCACCCCTGACCAACAGCAAGGTGGTGATCATC | 540  |
|                            | PN JP 2002542782-A/9   | QY  | 541  | GACGCCAACACAGCCGGGACCGGTGGTGAGCACAGTTCACCGTCTGCAACCGCGACGTGCTG  | 600  |
|                            | PD 17-DEC-2002   | Db  | 541  | GACGCCAACACAGCCGGGACCGGTGGTGAGCACAGTTCACCGTCTGCAACCGCGACGTGCTG  | 600  |
|                            | PF 20-APR-2000 JP 2000614390   | QY  | 601  | TGCATCTCCAGCATCCCCCGCCAGCGACAGCGACTACCCCTCCCGGGGAGATGTTCCCTG    | 660  |
|                            | PR 23-APR-1999 US 60/130694, 23-JUN-1999 US 60/140580 PI                 | Db  | 601  | TGCATCTCCAGCATCCCCCGCCAGCGACAGCGACTACCCCTCCCGGGGAGATGTTCCCTG    | 660  |
|                            | JENNIFER L HILLMAN, OLGA BANDMAN, TOM Y TANG, PREETI LAL, HENRY PI       | QY  | 661  | GACAGCGACGTGAACCCAGAGAACCCGCGGCAGATGGCGTGGTGGCCGGTATCACCCCTG    | 720  |
|                            | YUE, ROOPA REDDY, YALDA AZIMZAI, MARIAH R BAUGHN                         | Db  | 661  | GACAGCGACGTGAACCCAGAGAACCCGCGGCAGATGGCGTGGTGGCCGGTATCACCCCTG    | 720  |
|                            | PC C12N15/09, A61K31/711, A61K38/00, A61K45/00, A61K48/00, A61P1/16,     | QY  | 721  | GTGGGCTGTGCCACCCGCTGCAACGTGCCCGCGAGCAACTGCTCCTCCCGAGGGGACACC    | 780  |
|                            | PC A61P1/18,   | Db  | 721  | GTGGGCTGTGCCACCCGCTGCAACGTGCCCGCGAGCAACTGCTCCTCCCGAGGGGACACC    | 780  |
|                            | A61P3/10, A61P5/02, A61P5/06, A61P5/14, A61P5/26, A61P5/30, A61P7/00, PC | QY  | 781  | CCAGTGTAGACAAAGGGGACGGGGAGGTGGCCACCATCGCCAAACGGGAAGGTCAACCCCG   | 840  |
|                            | A61P7/02,  | Db  | 781  | CCAGTGTAGACAAAGGGGACGGGGAGGTGGCCACCATCGCCAAACGGGAAGGTCAACCCCG   | 840  |
|                            | PC A61P9/00, A61P9/10, A61P9/12, A61P13/00, A61P13/02, A61P13/ PC        | QY  | 841  | TCCAGTCCACAGAGGAGGCCACAGAGGCCACGAGAGTGCACAGACCCCTGGGCCCAGCGGAG  | 900  |
|                            | 08,  | Db  | 841  | TCCAGTCCACAGAGGAGGCCACAGAGGCCACGAGAGTGCACAGACCCCTGGGCCCAGCGGAG  | 900  |
|                            | PC A61P13/12, A61P15/00, A61P15/08, A61P15/12, A61P15/14, A61P17/00,     | QY  | 901  | CCAGAGACAGCCACATTCGGGCCCGGGCCCTCTCACAGAGCACGCTTCTCACTGACCCAGCC  | 960  |
|                            | PC A61P17/06,  | Db  | 901  | CCAGAGACAGCCACATTCGGGCCCGGGCCCTCTCACAGAGCACGCTTCTCACTGACCCAGCC  | 960  |
|                            | PC A61P19/00, A61P19/08, A61P19/10, A61P21/04, A61P25/08, A61P25/18,     | QY  | 961  | CCGACCCCGTCTCTTGCCCCCAGCCCTGGCAGCAGAAACGGGCGAGAGCCTGACAGCAGC    | 1020 |
|                            | PC A61P27/00,  | Db  | 961  | CCGACCCCGTCTCTTGCCCCCAGCCCTGGCAGCAGAAACGGGCGAGAGCCTGACAGCAGC    | 1020 |
|                            | PC A61P27/06, A61P27/12, A61P31/00, A61P35/00, A61P35/02, A61P37/02,     | QY  | 1021 | AGCACACGGCCAGAGCCAGAGCCAGCCAGCGGGACCCACGGGAGCAGGCAGTGTGCA       | 1080 |
|                            | PC A61P43/00,  | Db  | 1021 | AGCACACGGCCAGAGCCAGAGCCAGCCAGCGGGACCCACGGGAGCAGGCAGTGTGCA       | 1080 |
|                            | A61P43/00, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC | QY  | 1081 | CCACCATGTGCTGGGAGCCAGAAACGGCTGGCTCTATGTGCACTCGGCTGTGGCCAAC      | 1140 |
|                            | 10,  | Db  | 1081 | CCACCATGTGCTGGGAGCCAGAAACGGCTGGCTCTATGTGCACTCGGCTGTGGCCAAC      | 1140 |
|                            | PC C12P21/02, C12Q1/68, C12N15/00, C12N5/00, A61K37/02 CC Incyte ID      | QY  | 1141 | TGGAAGAAGTGCCTGCACCTCCATCAAGCTGAAGATTCTGTGTGAGCCCTGGTGCATGTC    | 1200 |
|                            | No.: 5733930CB1  | Db  | 1141 | TGGAAGAAGTGCCTGCACCTCCATCAAGCTGAAGATTCTGTGTGAGCCCTGGTGCATGTC    | 1200 |
|                            | FH Key Location/Qualifiers   | QY  | 1201 | AAAGSCCGTGTGCTGGTGGCTCTGGCGGACGGGACCCCTGGCCATCTTCCACCCGTGGTGAA  | 1260 |
|                            | FT source 1. .3694 /organism='Homo sapiens (human)'                      | Db  | 1201 | AAAGSCCGTGTGCTGGTGGCTCTGGCGGACGGGACCCCTGGCCATCTTCCACCCGTGGTGAA  | 1260 |
|                            | FT Location/Qualifiers   | QY  | 1261 | GATGCCAGTGGGATCTGAGCAACTATCACCTAATGGACCTGGGCCACCCGCAACCACTCC    | 1320 |
|                            | 1. .3694 /organism="Homo sapiens"  | Db  | 1261 | GATGCCAGTGGGATCTGAGCAACTATCACCTAATGGACCTGGGCCACCCGCAACCACTCC    | 1320 |
|                            | /mol_type="genomic DNA"  | QY  | 1321 | ATCCGCTGCATGGCTGTGTGTACGACCCGCTGTGGTGTGGCTACAGAAACAAGGTGCAC     | 1380 |
|                            | /db_xref="taxon:9606"  | Db  | 1321 | ATCCGCTGCATGGCTGTGTGTACGACCCGCTGTGGTGTGGCTACAGAAACAAGGTGCAC     | 1380 |
|                            |  | QY  | 1381 | GTATCCAGCCCCAAGACCATGCAGATAGAGAAGTCAATTTGACGCCACCCCGCGGGGAG     | 1440 |
|                            |  | Db  | 1381 | GTATCCAGCCCCAAGACCATGCAGATAGAGAAGTCAATTTGACGCCACCCCGCGGGGAG     | 1440 |
|                            |  | QY  | 1441 | AGCCAGGTGCGGACGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCTGGAC           | 1500 |
|                            |  | Db  | 1441 | AGCCAGGTGCGGACGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCTGGAC           | 1500 |
|                            |  | QY  | 1501 | TCCACCTGAGGCTTACCATGCACACGCAACAGCATCTACAGGACGTGGACATTGAG        | 1560 |
|                            |  | Db  | 1501 | TCCACCTGAGGCTTACCATGCACACGCAACAGCATCTACAGGACGTGGACATTGAG        | 1560 |
| FEATURES                   |  |     |      |   |      |
| source                     |  |     |      |   |      |
| ORIGIN                     |  |     |      |   |      |
| Query Match                | 100.0%; Score 3694; DB 6; Length 3694;                                   |     |      |   |      |
| Best Local Similarity      | 100.0%; Pred. No. 0;   |     |      |   |      |
| Matches 3694; Conservative | 0; Mismatches 0; Indels 0; Gaps 0;                                       |     |      |   |      |
| QY                         | 1 CTGGCAGGTTCTCTAGTGAGCCTCCTGGGCAGGGTCTGGAGGGTCGGAGAAGGAGGCCCGG          | 60  |      |   |      |
| Db                         | 1 CTGGCAGGTTCTCTAGTGAGCCTCCTGGGCAGGGTCTGGAGGGTCGGAGAAGGAGGCCCGG          | 60  |      |   |      |
| QY                         | 61 CGGCTCCACGCCCCACCTGCCTCAGTCTTGAAAGTGGCTTGCCGCGCAGAAAGCTGGCA           | 120 |      |   |      |
| Db                         | 61 CGGCTCCACGCCCCACCTGCCTCAGTCTTGAAAGTGGCTTGCCGCGCAGAAAGCTGGCA           | 120 |      |   |      |
| QY                         | 121 GCCCTACCTGTCTCAGATTTCTGGAGGATGGGTAGAGCCAGGGTTCGTGCCCCAGCGG           | 180 |      |   |      |
| Db                         | 121 GCCCTACCTGTCTCAGATTTCTGGAGGATGGGTAGAGCCAGGGTTCGTGCCCCAGCGG           | 180 |      |   |      |
| QY                         | 181 CCTCCCTGCTCCCTGCAGCTGAGTCCCAACCGGGGCGAGGAGACACGCGGATGAAGAAC          | 240 |      |   |      |
| Db                         | 181 CCTCCCTGCTCCCTGCAGCTGAGTCCCAACCGGGGCGAGGAGACACGCGGATGAAGAAC          | 240 |      |   |      |
| QY                         | 241 GTGCCGGTGCCTGACTGCGCCCTCTGGTGGAGAAGGACCCACCATGAAGCTGTGG              | 300 |      |   |      |
| Db                         | 241 GTGCCGGTGCCTGACTGCGCCCTCTGGTGGAGAAGGACCCACCATGAAGCTGTGG              | 300 |      |   |      |
| QY                         | 301 TGTGCCGGGGCGTCAACCTGAGCGGGTGGAGGCCCAATGAGGACGACGCTGGGAATGGA          | 360 |      |   |      |
| Db                         | 301 TGTGCCGGGGCGTCAACCTGAGCGGGTGGAGGCCCAATGAGGACGACGCTGGGAATGGA          | 360 |      |   |      |
| QY                         | 361 GTCAAGCGCGGCGTCAACCTGAGCGGGTGGAGGCCCAATGAGGACGACGCTGGGAATGGA         | 420 |      |   |      |
| Db                         | 361 GTCAAGCGCGGCGTCAACCTGAGCGGGTGGAGGCCCAATGAGGACGACGCTGGGAATGGA         | 420 |      |   |      |
| QY                         | 421 AAGAGCGCCACACGTCCTCCGAGAGAAAGAGGCCCAAGGAGCTCCCTGAAATGGACGCC          | 480 |      |   |      |

|    |      |  |      |
|----|------|--|------|
| QY | 1561 | CCCTACGTTCAGCAAGATGCTAGGCACCTGGCAAGCTGGGTTTCTCCTTCGTACGCATCACG | 1620 |
| Db | 1561 |  |      |
| QY | 1621 | CCCTACGTTCAGCAAGATGCTAGGCACCTGGCAAGCTGGGTTTCTCCTTCGTACGCATCACG | 1680 |
| Db | 1621 |  |      |
| QY | 1681 | GCCCTGCTTGTTCGCGGCAGCCGGCTCTGGGTGGGCACCGGCAACGGAGTGGTCACTCTCC  | 1740 |
| Db | 1681 |  |      |
| QY | 1741 | ATCCCCCTGACAGAGACTGTGGTCTTGACACCGAGGCCAGCTCCTGGGGCTCCGAGCCAAAT | 1800 |
| Db | 1741 |  |      |
| QY | 1801 | ATCCCCCTGACAGAGACTGTGGTCTTGACACCGAGGCCAGCTCCTGGGGCTCCGAGCCAAAT | 1860 |
| Db | 1801 |  |      |
| QY | 1861 | AAGACATCCCCCACCCTCTGGGGAGGGCGCCCGTCCCGGGGGCATCATCCACGTGTATGGC  | 1920 |
| Db | 1861 |  |      |
| QY | 1921 | AAGACATCCCCCACCCTCTGGGGAGGGCGCCCGTCCCGGGGGCATCATCCACGTGTATGGC  | 1980 |
| Db | 1921 |  |      |
| QY | 2041 | GATGACAGCAGTGACAGGGCGGCCAGCAGCTTTCATCCCTACTGCTCCATGGCCCCAGGCC  | 2100 |
| Db | 2041 |  |      |
| QY | 2101 | GATGACAGCAGTGACAGGGCGGCCAGCAGCTTTCATCCCTACTGCTCCATGGCCCCAGGCC  | 2160 |
| Db | 2101 |  |      |
| QY | 2161 | CAGCTATGCTTCCATGGGCACCCGGATGCCGTGAAGTTCTTTGTCTCGGTGCCAGGGAAC   | 2220 |
| Db | 2161 |  |      |
| QY | 2221 | CAGCTATGCTTCCATGGGCACCCGGATGCCGTGAAGTTCTTTGTCTCGGTGCCAGGGAAC   | 2280 |
| Db | 2221 |  |      |
| QY | 2281 | GTGCTGGCCACCCTGAATGGCAGTGTGCTGGACACCCAGCCAGCGGCCCCCTGGGCCAGCT  | 2340 |
| Db | 2281 |  |      |
| QY | 2341 | GTGCTGGCCACCCTGAATGGCAGTGTGCTGGACACCCAGCCAGCGGCCCCCTGGGCCAGCT  | 2400 |
| Db | 2341 |  |      |
| QY | 2401 | GCCCTGCTCGGAGCTGAGCGCCGAGTGAAGCTGCTGCCAACCTGCTGGTGTGAGCGGCGGG  | 2460 |
| Db | 2401 |  |      |
| QY | 2461 | GCCCTGCTCGGAGCTGAGCGCCGAGTGAAGCTGCTGCCAACCTGCTGGTGTGAGCGGCGGG  | 2520 |
| Db | 2461 |  |      |
| QY | 2521 | TGGGACATCGGATCAGCTGGGAGGAGGGGGTGTCTTCCACCCGAGGGGAA             | 2580 |
| Db | 2521 |  |      |
| QY | 2581 | AGTGGGGGATCGGATCAGCTGGGAGGAGGGGGTGTCTTCCACCCGAGGGGAA           | 2640 |
| Db | 2581 |  |      |

|    |      |   |      |
|----|------|---|------|
| QY | 2641 | GCCTAGGGAGCTGGGCCAGGCACTAGCCTTTGGCCAGGGAGGTGGGCCCTCAGGCTGCCCC | 2700 |
| Db | 2641 |   | 2700 |
| QY | 2701 | AGGTGCCCTGCACCCAGCCGCGCTTCTCTGGGGCTCCCGTCGTCAAGCCTCTATCCTG    | 2760 |
| Db | 2701 |   | 2760 |
| QY | 2761 | TCTGTCCCAACCCAGCTGTCCCTGCCCCAGGAGCTGGCATATAAAGCACGAGGCCGGC    | 2820 |
| Db | 2761 |   | 2820 |
| QY | 2821 | TCCCTGGGGAGCTGCTTGAGAAACAGAGACTGCTACCCCATCTCTGCCCATGCAGCAGGC  | 2880 |
| Db | 2821 |   | 2880 |
| QY | 2881 | TCTTGCCAGCCCCGTTCTGACCCCGTGTCCCCAGGCTCTGCCTGGGCAGAACTCACC     | 2940 |
| Db | 2881 |   | 2940 |
| QY | 2941 | TTGGAGGAGTGGGCCCCCTGAGTCTCTGTCCTCCCAAGAGCCCCCAGGTGGATTTCTCA   | 3000 |
| Db | 2941 |   | 3000 |
| QY | 3001 | GGCTGCCAGGGCAGGCCAGGCCTCAGGAAGAAGGGAGGCCCTTGGCCTCTCCGGGATC    | 3060 |
| Db | 3001 |   | 3060 |
| QY | 3061 | AGTCTTAGGACACAGGCTCAGCCTCAGGTTGATGGGGATGATGTGCTCCGGGGCCTGC    | 3120 |
| Db | 3061 |   | 3120 |
| QY | 3121 | CTCCTGCACGGGGCTCCACGGAGCCAGCTCCAGACACGCTACTAAGTGCCTAGGGTTG    | 3180 |
| Db | 3121 |   | 3180 |
| QY | 3181 | CCCGCTGTGGCCTGCTCCAGGGAGCAACAGAGAGGCCACCAAGCAGAGGCCCGTGGGC    | 3240 |
| Db | 3181 |   | 3240 |
| QY | 3241 | TGAGGATGGAGCCGCCCCAGCCGACTCCAAGCCCGCAGAGGGCAGACGCCACCCTGGAC   | 3300 |
| Db | 3241 |   | 3300 |
| QY | 3301 | TGCTCTCCCTGCCAGCTGGGCTCTCTGGCCTATTCTACCTTCCAGGCCCACTGCAC      | 3360 |
| Db | 3301 |   | 3360 |
| QY | 3361 | CCTGTCTGGAGGCCCTTATGAGGCGAGCCAGCCCGCACCCCGCACCCCAACCCAGAGAA   | 3420 |
| Db | 3361 |   | 3420 |
| QY | 3421 | GCACAGATCTTGGGAGCTGCCCAAGCCCGCTGGCCACCCGAGGCTGCAGCCGCTG       | 3480 |
| Db | 3421 |   | 3480 |
| QY | 3481 | CGTGCCGGCTTCTCCCAACCACTGCGCCACTCCACTGTGATGATGTCTCCGCTCCCTCG   | 3540 |
| Db | 3481 |   | 3540 |
| QY | 3541 | TCTGTTCCCCAGGATCTCGAAGTGACTCCGGGCTGAGCAGTGGGGCGCTGGGGAGGG     | 3600 |
| Db | 3541 |   | 3600 |
| QY | 3601 | GTGACGATTCTCCTCAGGCTTTGGCCCTGCAAGCAAAACCCACATATCTGTATGTA      | 3660 |
| Db | 3601 |   | 3660 |
| QY | 3661 | ATAAATGTCCTTAACGTCGTAAAAAATAAAAAA                             | 3694 |
| Db | 3661 |   | 3694 |

```
RESULT 2
AX041037
LOCUS      AX041037          3694 bp          DNA          linear          PAT 23-NOV-2000
DEFINITION Sequence 26 from Patent WO0065054.
ACCESSION  AX041037
VERSION     AX041037.1  GI:11340615
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Hillman,J.L., Bandman,O., Tang,Y.T., Lal,P., Yue,H., Redd,R.,
            Azimzai,Y. and Baughn,M.R.
TITLE       Human membrane-associated proteins
JOURNAL     Patent: WO 0065054-A 26 02-NOV-2000;
            Incyte Genomics, Inc. (US)
FEATURES
source      Location/Qualifiers
            1..3694
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
            /note="Incyte ID No.: 5733930CB1"

ORIGIN
Query Match      100.0%; Score 3694; DB 6; Length 3694;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3694; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  CTGGCAGGTTCCCTAGTGAGCCCTCTGGGCAGGGTCTGGAGGGTGGAGAGGAGGAGCCCG 60
Db      1  CTGGCAGGTTCCCTAGTGAGCCCTCTGGGCAGGGTCTGGAGGGTGGAGAGGAGGAGCCCG 60

QY      61  CGGCTCCACGCCCCACCTGCCCTCAGTCTTGAAGTGGCTCTGCCGCCAAGAAGTGGCA 120
Db      61  CGGCTCCACGCCCCACCTGCCCTCAGTCTTGAAGTGGCTCTGCCGCCAAGAAGTGGCA 120

QY      121  GCCCTACCTGTCTCAGATTCTGGAGGGATGGGTAGGAGCCAGGGTTCGTGCCACGGCG 180
Db      121  GCCCTACCTGTCTCAGATTCTGGAGGGATGGGTAGGAGCCAGGGTTCGTGCCACGGCG 180

QY      181  CCTCCCTGCTCCCTGCAGCTGAGTCCCAACGGGGGCAGGAGGACACGGCGATGAAGAAC 240
Db      181  CCTCCCTGCTCCCTGCAGCTGAGTCCCAACGGGGGCAGGAGGACACGGCGATGAAGAAC 240

QY      241  GTGCGGGTGCCGGTGTACTGCGCCCTCTGGTGGAGAGGACCCACCATGAAGCTGTGG 300
Db      241  GTGCGGGTGCCGGTGTACTGCGCCCTCTGGTGGAGAGGACCCACCATGAAGCTGTGG 300

QY      301  TGTGCGCGGGCGTCAACCTGAGCGGGTGGAGGCCCAATGAGGACGACGCTGGGAATGGA 360
Db      301  TGTGCGCGGGCGTCAACCTGAGCGGGTGGAGGCCCAATGAGGACGACGCTGGGAATGGA 360

QY      361  GTCAAGCCAGCGCCAGCGCGCGATCCCTGACCTGCGACCGCGAAGGAGACGGCGAGCCC 420
Db      361  GTCAAGCCAGCGCCAGCGCGCGATCCCTGACCTGCGACCGCGAAGGAGACGGCGAGCCC 420

QY      421  AAGAGCGCCACACGCTCTCCCGAGAGAAGAGCCAGGAGCTCCCTGAAAATGGACGCC 480
Db      421  AAGAGCGCCACACGCTCTCCCGAGAGAAGAGCCAGGAGCTCCCTGAAAATGGACGCC 480

QY      481  ACCTCCAGCCGGGTGTGGATCCTGACGAGCACCTGACCCACCCAGCAAGGTGGTATCATC 540
Db      481  ACCTCCAGCCGGGTGTGGATCCTGACGAGCACCTGACCCACCCAGCAAGGTGGTATCATC 540

QY      541  GACGCCAACACGCGGGCAGCGGTGGTGAACAGTTCACCGTCTGCAACCGCGCACGTGCTG 600
Db      541  GACGCCAACACGCGGGCAGCGGTGGTGGACCAAGTTCACCGTCTGCAACCGCGCACGTGCTG 600

QY      601  TGCATCTCCAGCATCCCGCGGGCCAGCGACAGCGACTACCCCTCCCGGGGAGATGTTCTCTG 660
Db      601  TGCATCTCCAGCATCCCGCGGGCCAGCGACAGCGACTACCCCTCCCGGGGAGATGTTCTCTG 660
```

```
661  GACAGCGACGTGAACCCAGAGGACCCGGGGCAGATGGCGTGCTGGCCGGTATCACCCCTG 720
661  GACAGCGACGTGAACCCAGAGGACCCGGGGCAGATGGCGTGCTGGCCGGTATCACCCCTG 720
721  GTGGGCTGTGCCACCCGCTGCAACGTCGCGGGAGCAACTGCTCCTCCGAGGGGACACC 780
721  GTGGGCTGTGCCACCCGCTGCAACGTCGCGGGAGCAACTGCTCCTCCGAGGGGACACC 780
781  CCAGTGCTAGACAAGGGGACGGGGAGGTGGCCACCATCGCCAAACGGGAAGGTCAACCCG 840
781  CCAGTGCTAGACAAGGGGACGGGGAGGTGGCCACCATCGCCAAACGGGAAGGTCAACCCG 840
841  TCCCAGTCCACAGAGGAGGCCACAGAGGCCACGAGGAGTGCACACCCCTGGGCCCGAGG 900
841  TCCCAGTCCACAGAGGAGGCCACAGAGGCCACGAGGAGTGCACACCCCTGGGCCCGAGG 900
901  CCAGAGACAGCCACATTGCGGCCCGGGCCTCTCACAGAGCAGTCTTCACTGACCCAGCC 960
901  CCAGAGACAGCCACATTGCGGCCCGGGCCTCTCACAGAGCAGTCTTCACTGACCCAGCC 960
961  CCGACCCCGTCTCTGGCCCCCAGCCTGGCAGCGGAAACGGGGCAGAGCCTGACAGCAGC 1020
961  CCGACCCCGTCTCTGGCCCCCAGCCTGGCAGCGGAAACGGGGCAGAGCCTGACAGCAGC 1020
1021  AGCACACGGCCAGAGCCAGAGCCAGCGGGAGACCCACCGGAGCAGGACGAGTGTCTGCA 1080
1021  AGCACACGGCCAGAGCCAGAGCCAGCGGGAGACCCACCGGAGCAGGACGAGTGTCTGCA 1080
1081  CCCACCATGTGGCTGGGAGCCAGAAACGGCTGGCTCTATGTGCACTCGGCTGTGGCCAA 1140
1081  CCCACCATGTGGCTGGGAGCCAGAAACGGCTGGCTCTATGTGCACTCGGCTGTGGCCAA 1140
1141  TGGAAAGAGTGCCTGCACCTCCATCAAGCTGAAGATTCTGTGTGAGCCCTGGTGCATGTC 1200
1141  TGGAAAGAGTGCCTGCACCTCCATCAAGCTGAAGATTCTGTGTGAGCCCTGGTGCATGTC 1200
1201  AAAGGCCGTGTGTGGTGGCTCTGGCGGACGGGACCCCTGGCCATCTTCCACCGTGGTGA 1260
1201  AAAGGCCGTGTGTGGTGGCTCTGGCGGACGGGACCCCTGGCCATCTTCCACCGTGGTGA 1260
1261  GATGGCCAGTGGGATCTGAGCAACTATCACCTAATGGACCTGGGCCACCCGACCACTCC 1320
1261  GATGGCCAGTGGGATCTGAGCAACTATCACCTAATGGACCTGGGCCACCCGACCACTCC 1320
1321  ATCCGCTGCATGGCTGTGTGTACGACCGCGTGTGGTGTGGCTACAAGAACAAAGTGCAC 1380
1321  ATCCGCTGCATGGCTGTGTGTACGACCGCGTGTGGTGTGGCTACAAGAACAAAGTGCAC 1380
1381  GTCAATCCAGCCCAAGACCATGAGATAGAGAAAGTCAATTTGACGCCCAACCCGCGGGAG 1440
1381  GTCAATCCAGCCCAAGACCATGAGATAGAGAAAGTCAATTTGACGCCCAACCCGCGGGAG 1440
1441  AGCCAGGTGCGGCAGCTGGCGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCTGGAC 1500
1441  AGCCAGGTGCGGCAGCTGGCGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCTGGAC 1500
1501  TCCACCTGAGGCTTACCATGCACACGACCAAGCATCTACAGGACGTGGACATTGAG 1560
1501  TCCACCTGAGGCTTACCATGCACACGACCAAGCATCTACAGGACGTGGACATTGAG 1560
1561  CCCTACGTCAGCAAGATCTAGGCACTGGCAAGCTGGGTTTCTCCTTCGTACGCATCAGC 1620
1561  CCCTACGTCAGCAAGATCTAGGCACTGGCAAGCTGGGTTTCTCCTTCGTACGCATCAGC 1620
1621  GCCCTGCTTGTGCGGGCAGCCGGCTCTGGGTGGGCAACGGCAACGGAGTGGTCACTCC 1680
1621  GCCCTGCTTGTGCGGGCAGCCGGCTCTGGGTGGGCAACGGCAACGGAGTGGTCACTCC 1680
1681  ATCCCCCTGACAGAGACTGTGGTCTCTGCACCGAGGCCAGCTCCTGGGGCTCCGAGCCAT 1740
1681  ATCCCCCTGACAGAGACTGTGGTCTCTGCACCGAGGCCAGCTCCTGGGGCTCCGAGCCAT 1740
1741  AAGACATCCCCCACTCTGGGGAGGGCGCCCGTCCCGGGGGCATCATCCAGTGTATGGC 1800
```





Db 3123 AGGGACATGAGCCAGGTGAAGCCCGTGTGTCCAAAGGCAGAGCGCAGTCACATCATCGT 3182

QY 2160 GTGGCAGGTGTCTTACACCCCGAGTGAAGCTGCTGCTCCCTGCTCCCTGGCCCCGACCTGTACAT 2219

Db 3183 GTGGCAGGTGTCTTACACCCCGAGTGAAGCTGCTGCTCCCTGCTCCCTGGCCCCGACCTGTACAT 3242

QY 2220 AGGACCCCGACACCTGACCCCGCCCGCCCGCGGGGTAGCCAGCAGGCGCCGCCG 2279

Db 3243 AGGACCCCGACACCTGACCCCGCCCGCCCGCGGGGTAGCCAGCAGGCGCCGCCG 3302

QY 2280 CCCTCTTAAACCTCTCAACCTGCAGCTTTTCACTGAGTCTGGCCCCCTCCAGCGGGCAGG 2339

Db 3303 CCCTCTTAAACCTCTCAACCTGCAGCTTTTCACTGAGTCTGGCCCCCTCCAGCGGGCAGG 3362

QY 2340 GAGTCGGGGATGCGGATCAGCTGGGAGGAGGAGGGAGGGGTGCTTCCACCCGAGGGGA 2399

Db 3363 GAGTCGGGGATGCGGATCAGCTGGGAGGAGGAGGGAGGGGTGCTTCCACCCGAGGGGA 3422

QY 2400 AGATGCTCTCGGGACAGTTTCCCGGCAGCTCCTGGCCAGCTTCAGGCCCAGAGTCCTCA 2459

Db 3423 AGATGCTCTCGGGACAGTTTCCCGGCAGCTCCTGGCCAGCTTCAGGCCCAGAGTCCTCA 3482

QY 2460 AGTCCAGGGCACCTTTGGGCCCAGCGCAGGCAGAAATCCGAGGTGTCCTGGCTCTACCCCTG 2519

Db 3483 AGTCCAGGGCACCTTTGGGCCCAGCGCAGGCAGAAATCCGAGGTGTCCTGGCTCTACCCCTG 3542

QY 2520 GGCCTCTACTCCCCAGCACCCCTGGAGGAGGAGGGGCTCCCCGCCCGCAGGCTGCCT 2579

Db 3543 GGCCTCTACTCCCCAGCACCCCTGGAGGAGGAGGGGCTCCCCGCCCGCAGGCTGCCT 3602

QY 2580 GCCCTGGGCCACCTCTGCATGTCTCATGGGGCCACCCTGCTCCTGGGCCCTCACTC 2639

Db 3603 GCCCTGGGCCACCTCTGCATGTCTCATGGGGCCACCCTGCTCCTGGGCCCTCACTC 3662

QY 2640 TGCTAGGGGAGCTGGGCCAGGCACTAGCCTTTGCCCAGGAGGTGGCCTCAGGCTGCC 2699

Db 3663 TGCTAGGGGAGCTGGGCCAGGCACTAGCCTTTGCCCAGGAGGTGGCCTCAGGCTGCC 3722

QY 2700 CAGGTGCTTGACCCCCAGCCCGCTTCTCTGGGGCCTCCCGTCGTCAAGCCTCTATCCT 2759

Db 3723 CAGGTGCTTGACCCCCAGCCCGCTTCTCTGGGGCCTCCCGTCGTCAAGCCTCTATCCT 3782

QY 2760 GTCTGTCCCCACCCACGCTGTCCCCTGCCCAGGAGCTGGCATAAAGCACAGGCCCCGG 2819

Db 3783 GTCTGTCCCCACCCACGCTGTCCCCTGCCCAGGAGCTGGCATAAAGCACAGGCCCCGG 3842

QY 2820 CTCCTGGGAGCTGCTTGAGAACAGAGACTGTACCCCATCCTGCCCCATGCAGGCAGG 2879

Db 3843 CTCCTGGGAGCTGCTTGAGAACAGAGACTGTACCCCATCCTGCCCCATGCAGGCAGG 3902

QY 2880 CTCTTGCCAGCCCCGTTCTGACCCGCTGTCCCCCAGGCTCTGCTGGGCGAGGAGCTCAC 2939

Db 3903 CTCTTGCCAGCCCCGTTCTGACCCGCTGTCCCCCAGGCTCTGCTGGGCGAGGAGCTCAC 3962

QY 2940 CTTGGAGGAGTGGGCCCTTGAGTCTGTCCCTCCAGAAAGCCCCCAGGGTGGGATTTCTC 2999

Db 3963 CTTGGAGGAGTGGGCCCTTGAGTCTGTCCCTCCAGAAAGCCCCCAGGGTGGGATTTCTC 4022

QY 3000 AGGTGCCAGGGCAGGCCCTCAGGAAGAAGGGAGGCCCTGTCCTGGCGAGGAGCTCAC 3059

Db 4023 AGGTGCCAGGGCAGGCCCTCAGGAAGAAGGGAGGCCCTGTCCTGGCGAGGAGCTCAC 4082

QY 3060 CAGTCTAGGACACAGGCTCAGCCTCAGGTGATGGGGATGATGTCTCCCGGGCCTG 3119

Db 4083 CAGTCTAGGACACAGGCTCAGCCTCAGGTGATGGGGATGATGTCTCCCGGGCCTG 4142

QY 3120 CCTCTGACGGGGCTCCAGGAGCCCCAGCTCCAGACACGCTACTAAGTGCCTAGGGTT 3179

Db 4143 CCTCTGACGGGGCTCCAGGAGCCCCAGCTCCAGACACGCTACTAAGTGCCTAGGGTT 4202

QY 3180 GCCCGCTGTGGCCTGTCTCCAGGGAGCAACAGAGAGGCCACCAAGCAGAGGCCCTGGGG 3239

Db 4203 GCCCGCTGTGGCCTGTCTCCAGGGAGCAACAGAGAGGCCACCAAGCAGAGGCCCTGGGG 4262

QY 3240 CTGAGGATGGAGCCCGCCCCCAGCCGACTCCAAGCCGAGAGGGCAGAGCCACCCCTGGA 3299

Db 4263 CTGAGGATGGAGCCCGCCCCCAGCCGACTCCAAGCCGAGAGGGCAGAGCCACCCCTGGA 4322

QY 3300 CTGCTCTCCCTGCCCAGCTGGGCCCTCTCTGGCCCTATTCTTCCCTCAGGCCCACTGCAC 3359

Db 4323 CTGCTCTCCCTGCCCAGCTGGGCCCTCTCTGGCCCTATTCTTCCCTCAGGCCCACTGCAC 4382

QY 3360 TCCTGTCTGGGAGGCCCTTATGAGGCGAGCCCGCCCGCACCACCCCAACCAAGAGA 3419

Db 4383 TCCTGTCTGGGAGGCCCTTATGAGGCGAGCCCGCCCGCACCACCCCAACCAAGAGA 4442

QY 3420 AGCACAGATCTTGGGGAGCTGCCCAACAAGCCCGCTGGCCAACGAGGGCTGCAGCGCT 3479

Db 4443 AGCACAGATCTTGGGGAGCTGCCCAACAAGCCCGCTGGCCAACGAGGGCTGCAGCGCT 4502

QY 3480 GCGCTGCCGGCTTCTCCCCACCAACCTGCCACCTCCACTGTGATGTATGTCCGCTCCCTC 3539

Db 4503 GCGCTGCCGGCTTCTCCCCACCAACCTGCCACCTCCACTGTGATGTATGTCCGCTCCCTC 4562

QY 3540 GTCTGTTCCCCCAGGATCTCGAAGTGACTCCGGGCTGAGCAGTGGGGCGGCTGGGGGAGG 3599

Db 4563 GTCTGTTCCCCCAGGATCTCGAAGTGACTCCGGGCTGAGCAGTGGGGCGGCTGGGGGAGG 4622

QY 3600 GGTGACGATTTCTCTCAGGCTTGGCCCTGCAAGCAAAACCCACATATCTGCTCTGTATGT 3659

Db 4623 GGTGACGATTTCTCTCAGGCTTGGCCCTGCAAGCAAAACCCACATATCTGCTCTGTATGT 4682

QY 3660 AATAAATGTCTTAACGTCTG 3678

Db 4683 AATAAATGTCTTAACGTCTG 4701

RESULT 4

AB028989

LOCUS AB028989 5621 bp mRNA linear PRI 07-JAN-2003

DEFINITION Homo sapiens mRNA for KIAA1066 protein, partial cds.

ACCESSION AB028989

VERSION AB028989.2 GI:27529813

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Kikuno,R., Nagase,T., Ishikawa,K., Hirose,M., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.

TITLE Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro

JOURNAL DNA Res. 6 (3), 197-205 (1999)

MEDLINE 99397452

PUBMED 10470851

REFERENCE 2 (bases 1 to 5621)

AUTHORS Ohara,O., Nagase,T. and Kikuno,R.

TITLE Direct Submission

JOURNAL Submitted (17-JUN-1999) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:81-438-52-3913, Fax:81-438-52-3914)

COMMENT On Jan 7, 2003 this sequence version replaced gi:5689468.

FEATURES

source

1. .5621

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/chromosome="16"

/clone="ah03682"

/tissue type="brain"

/clone\_lib="pBluescriptII SK plus"

1. .5621

/gene="KIAA1066"

gene

CDS

```
<91..4131
/gene="KIAA1066"
/note="This sequence was replaced that of hj05363b cDNA as
a representative cDNA sequence for KIAA1066."
/codon_start=1
/product="KIAA1066 protein"
/protein_id="BAAB3018.2"
/db_xref="GI:27529814"
/translation="RAAALAAVAAMMEIQMDEGGVVVYQDDYCSGWSMSEVSGLAG
SIYREFERLIHCYDEEVVKELMPLVNVLENLDSVLSENQEHVELELLREDNEQLLT
QYERERALRRQAEKEFTFEFEDALEQEKELQIOVEHYEFQTRQLELKAKNYADQISRL
EERESMKKEYNALHQHTEMIQTYVEHIERSKMQQVGGNSQTESLPGRRKERPTSL
NVFPLADGTVRAQIGGKLVPA GDHWHLSLDLQLOQSSSYQCPODEMSESGSSAAATP
STTGKSNTPSTSSVPSAAVTPLNESLQPLGDYGVGSKNSKRAREKRDSRNMEVQVTOE
MRNVSIGMSSDEWSDVDQIIDSTPELDMCPETRLDRGTSSPTQGIYNKAFGINTDSL
YHELSTAGSEVIGDVDEGADLLGEFVRDQDFGMGKEVGNLLENLSQLLETKNALNVV
KNDLIAKVDQLSGEQEVLRLGELEAAKQAKVKLENRIKELEELKRVKSEAI IARREP
KEEDVSSYLCITESDKTPMAQRRRTFRVEMARVLMERNQYKERLMEIQEAVRWTEMLR
ASREHPSVQEKKSTIQOFFSRLFSSSSPPPAKRPVPSVNIHYKSPPTAGFSQRRNH
AMCPI SAGSRPLEFFDDDTSSARREKREQYRQVREHVRNDDRLQACGWSLPAKY
KQLSPNGQEDTRMKNVPVPVYCRPLVEKQPTMKLWCAAGVNLSGWRPNEDDAGNGVK
PAPGRDPLTCDREGDGEPKSAHTSPEKKKAKELPEMDATSSRWIILSTLTTSKVII
DANQPGTVVDQFTVCNAHVLCISSIPAASDSYPPGEMFLDSDVNPEPDGADVLGI
TLVGCAATCNVPRNSCSRGTDPVLDKQGEVATIANGVNPSQSTERATEATEVPDP
GPSEPETATLRPGPLTEHVFTDPA TPSSGPPQPSGSENGPEPDSSTRPEPEPSGDTG
AGSSAAPNWLGAQNGWLYVSHAVANWKKLHSIKLXDSLVLVHVKRVLVALADGT
LAIFHRGEDQWDLNHYHMDLGHPHHSIRCMVAVYRVWCGYKKNKHVVIQPKTMQIE
KSFDAHPRESQVRQLAWITGDGVVWSIRLDSLRLYHATHQHLQVDJLEPYVSKMLG
TGKLGFSFVRITALLVAGSRLVGTGVGVVISIPLTETVVLHRLGQLLRANKTSPTS
GEGARPGGIHIVYGDSSDRAASSFI PYCSMAQAQLCFHHRDVAKFVSVPGNVLAT
LNGSVLDSPAEGPGPAAPASEVEGQKLRNLVLSGGEYIDFRIGDGEDDETEEGAGD
MSQVKPVLKSAERSHIIIVQVSYTPE"
```

ORIGIN

Query Match 94.3%; Score 3482; DB 9; Length 5621;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 3488; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

|    |      |  |      |
|----|------|--|------|
| QY | 180  | GCCTCCCTGCTCCCTGCAGCTGAGTCCCAACGGGGCCAGGAGGACACGCGGATGAAGAA    | 239  |
| Db | 2124 | GCCCGCCAAGTACAAGCAGCTGAGTCCCAACGGGGCCAGGAGGACACGCGGATGAAGAA    | 2183 |
| QY | 240  | CGTGCCGGTGCCGGTGTA CTCCGCTCTGCTGGTGGAGAAGGACCCCAACCATGAAGCTGTG | 299  |
| Db | 2184 | CGTGCCGGTGCCGGTGTA CTCCGCTCTGCTGGTGGAGAAGGACCCCAACCATGAAGCTGTG | 2243 |
| QY | 300  | GTGTGCCCGGGCGCTCAACCTGAGCGGGTGGAGGCCCAATGAGGACGACGCTGGGAATGG   | 359  |
| Db | 2244 | GTGTGCCCGGGCGCTCAACCTGAGCGGGTGGAGGCCCAATGAGGACGACGCTGGGAATGG   | 2303 |
| QY | 360  | AGTCAAGCCAGCGCCAGCGCGATCCCTTGACCTGCGACCCGGAAGGACGCGGAGCC       | 419  |
| Db | 2304 | AGTCAAGCCAGCGCCAGCGCGATCCCTTGACCTGCGACCCGGAAGGACGCGGAGCC       | 2363 |
| QY | 420  | CAAGAGCCGCCACACGCTCTCCCGAGAAGAAGAGGCCAAGGAGCTCCCTGAAATGGACGC   | 479  |
| Db | 2364 | CAAGAGCCGCCACACGCTCTCCCGAGAAGAAGAGGCCAAGGAGCTCCCTGAAATGGACGC   | 2423 |
| QY | 480  | CACCTCCAGCCGGGTGTGGATCTCTGACAGCACCCCTGACCACCCAGCAAGGTGGTGATCAT | 539  |
| Db | 2424 | CACCTCCAGCCGGGTGTGGATCTCTGACAGCACCCCTGACCACCCAGCAAGGTGGTGATCAT | 2483 |
| QY | 540  | CGACGCCAACACGACCGGGCACGGTGGTGGACCAAGTTTACCGTCTGCAACGCGCACGTGCT | 599  |
| Db | 2484 | CGACGCCAACACGACCGGGCACGGTGGTGGACCAAGTTTACCGTCTGCAACGCGCACGTGCT | 2543 |
| QY | 600  | GTGCATCTCCAGCATCCCCGGGGCCAGCGACAGCGACTACCTCCCGGGGAGATGTTCCCT   | 659  |
| Db | 2544 | GTGCATCTCCAGCATCCCCGGGGCCAGCGACAGCGACTACCTCCCGGGGAGATGTTCCCT   | 2603 |
| QY | 660  | GGACAGCGACCTGAACCCAGAGGACCCGGGCGCAGATGGCGTGTGGCCGCTATCACCCCT   | 719  |
| Db | 2604 | GGACAGCGACCTGAACCCAGAGGACCCGGGCGCAGATGGCGTGTGGCCGCTATCACCCCT   | 2663 |

|    |      |   |      |
|----|------|---|------|
| QY | 720  | GGTGGGCTGTGCCACCCCGCTGCAACAGTGC CGGGAGCAACTGCTCTCTCCCGAGGGACAC  | 779  |
| Db | 2664 | GGTGGGCTGTGCCACCCCGCTGCAACAGTGC CGGGAGCAACTGCTCTCTCCCGAGGGACAC  | 2723 |
| QY | 780  | CCCAGTGTAGACAAGGGGCGAGGGGAGGTGGGCACCATCGCCAAACGGGAAGTCAACCC     | 839  |
| Db | 2724 | CCCAGTGTAGACAAGGGGCGAGGGGAGGTGGGCACCATCGCCAAACGGGAAGTCAACCC     | 2783 |
| QY | 840  | GTCCCAGTCCACAGAGGAGGCCACAGAGGCCACGAGGTTGCCAGACCTTGGGCCAGCGA     | 899  |
| Db | 2784 | GTCCCAGTCCACAGAGGAGGCCACAGAGGCCACGAGGTTGCCAGACCTTGGGCCAGCGA     | 2843 |
| QY | 900  | GCCAGAGACAGCCACATTGCGGCCCGGGCTCTCACAGAGCACGTCTTCACTGACCCAGC     | 959  |
| Db | 2844 | GCCAGAGACAGCCACATTGCGGCCCGGGCTCTCACAGAGCACGTCTTCACTGACCCAGC     | 2903 |
| QY | 960  | CCCGACCCCGTCTTGGCCCCCAGCCCTGGCAGCGAGAACCGGGCAGAGCCTGACAGCAG     | 1019 |
| Db | 2904 | CCCGACCCCGTCTTGGCCCCCAGCCCTGGCAGCGAGAACCGGGCAGAGCCTGACAGCAG     | 2963 |
| QY | 1020 | CAGCACACGGCCAGAGCCAGAGCCAGCGGGGACCCACGGGAGAGGAGCAGCAGTGTGC      | 1079 |
| Db | 2964 | CAGCACACGGCCAGAGCCAGAGCCAGCGGGGACCCACGGGAGAGGAGCAGCAGTGTGC      | 3023 |
| QY | 1080 | ACCCACCATGTGGTGGGAGCCACAGAACGGCTGGCTCTATGTGCACTCGGCTGTGGCAA     | 1139 |
| Db | 3024 | ACCCACCATGTGGTGGGAGCCACAGAACGGCTGGCTCTATGTGCACTCGGCTGTGGCAA     | 3083 |
| QY | 1140 | CTGGAAGAAAGTGCCTGCACCTCCATCAAGCTGAAGGATTCTGTGCTGAGCCTGGTGATGT   | 1199 |
| Db | 3084 | CTGGAAGAAAGTGCCTGCACCTCCATCAAGCTGAAGGATTCTGTGCTGAGCCTGGTGATGT   | 3143 |
| QY | 1200 | CAAAGCCCGTGTGTGGTCTGTGGCGGACGGGACCTGGCCATCTTCCACCGTGGTGA        | 1259 |
| Db | 3144 | CAAAGCCCGTGTGTGGTCTGTGGCGGACGGGACCTGGCCATCTTCCACCGTGGTGA        | 3203 |
| QY | 1260 | AGATGGCCAGTGGGATCTGAGCAACTATCACCTAATGGACCTGGGCCACCCGACCACTC     | 1319 |
| Db | 3204 | AGATGGCCAGTGGGATCTGAGCAACTATCACCTAATGGACCTGGGCCACCCGACCACTC     | 3263 |
| QY | 1320 | CATCCGCTGCATGGCTGTGTGTACGACCGCGTGTGGTGTGGCTACAAGAAACAAGGTGCA    | 1379 |
| Db | 3264 | CATCCGCTGCATGGCTGTGTGTACGACCGCGTGTGGTGTGGCTACAAGAAACAAGGTGCA    | 3323 |
| QY | 1380 | CGTCATCCAGCCCAAGACCATGCAGATAGAGAAAGTCATTGTAGCCCAACCCGCGGCGGA    | 1439 |
| Db | 3324 | CGTCATCCAGCCCAAGACCATGCAGATAGAGAAAGTCATTGTAGCCCAACCCGCGGCGGA    | 3383 |
| QY | 1440 | GAGCAGGTGCGGCAGCTGGCGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCTTGA      | 1499 |
| Db | 3384 | GAGCAGGTGCGGCAGCTGGCGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCTTGA      | 3443 |
| QY | 1500 | CTCCACCTGAGGCTCTACCATGCACACACGACCACTGAGGACCTTACAGGACGTGGACATTGA | 1559 |
| Db | 3444 | CTCCACCTGAGGCTCTACCATGCACACACGACCACTGAGGACCTTACAGGACGTGGACATTGA | 3503 |
| QY | 1560 | GCCCTACGTGAGCAAGATGCTAGGCACTGGCAAGCTGGGTTTCTCCTTGTACGCATCAC     | 1619 |
| Db | 3504 | GCCCTACGTGAGCAAGATGCTAGGCACTGGCAAGCTGGGTTTCTCCTTGTACGCATCAC     | 3563 |
| QY | 1620 | GGCCCTGCTTGTGCGGGCAGCCGGCTCTGGGTGGGACCGGCAACCGGAGTGGTCACTC      | 1679 |
| Db | 3564 | GGCCCTGCTTGTGCGGGCAGCCGGCTCTGGGTGGGACCGGCAACCGGAGTGGTCACTC      | 3623 |
| QY | 1680 | CATCCCTCTGACAGAGACTGTGGTCTGACCGAGGCCAGCTCCTGGGCTCCGAGCCAA       | 1739 |
| Db | 3624 | CATCCCTCTGACAGAGACTGTGGTCTGACCGAGGCCAGCTCCTGGGCTCCGAGCCAA       | 3683 |
| QY | 1740 | TAAGACATCCCCCACCCTCTGGGGAGGGGCGCCGCTCCCGGGGCATCATCCACGTGTATGG   | 1799 |
| Db | 3684 | TAAGACATCCCCCACCCTCTGGGGAGGGGCGCCGCTCCCGGGGCATCATCCACGTGTATGG   | 3743 |

QY 1800 CGATGACAGCAGTGTACAGGGCGGCCAGCAGCTTTCATCCCTACTGCTCCTCATGGCCCGCAGGC 1859  
Db |||||||  
3744 CGATGACAGCAGTGTACAGGGCGGCCAGCAGCTTTCATCCCTACTGCTCCTCATGGCCCGCAGGC 3803  
QY 1860 CCAGCTATGCTTCCATGGGCACCGCGGATGCCGTGAAGTCTTTGTCTCGGTGCCAGGAA 1919  
Db |||||||  
3804 CCAGCTATGCTTCCATGGGCACCGCGGATGCCGTGAAGTCTTTGTCTCGGTGCCAGGAA 3863  
QY 1920 CGTGTGGCCACCTCGAATGGCAGTGTGCTGGACAGCCAGCCAGCGGCCCTTGGGCCAGC 1979  
Db |||||||  
3864 CGTGTGGCCACCTCGAATGGCAGTGTGCTGGACAGCCAGCCAGCGGCCCTTGGGCCAGC 3923  
QY 1980 TGCCCTGCTCGGAGGTTCGAGGCCAGAAAGCTGCCGAACGTGCTGTGCTGAGCGCGG 2039  
Db |||||||  
3924 TGCCCTGCTCGGAGGTTCGAGGCCAGAAAGCTGCCGAACGTGCTGTGCTGAGCGCGG 3983  
QY 2040 GGAGGGCTACATCGACTTCCGCAATTGGAGACCGAGAGGACGACGAGACGGAGGGCGC 2099  
Db |||||||  
3984 GGAGGGCTACATCGACTTCCGCAATTGGAGACCGAGAGGACGACGAGACGGAGGGCGC 4043  
QY 2100 AGGGGACATGAGCCAGGTGAAGCCCGTGTGTGCCAGGACGAGCGAGTCAATCATCGT 2159  
Db |||||||  
4044 AGGGGACATGAGCCAGGTGAAGCCCGTGTGTGCCAGGACGAGCGAGTCAATCATCGT 4103  
QY 2160 GTGGCAGGTGTCTTACACCCCGAGTGAAGCTGTGCTGCCCTGCCCGCCGAGTCAATCAT 2219  
Db |||||||  
4104 GTGGCAGGTGTCTTACACCCCGAGTGAAGCTGTGCTGCCCTGCCCGCCGAGTCAATCAT 4163  
QY 2220 AGGACCCCGCAGCACCTGACCCCGCCCGCGGGGTAGCCAGCCAGCGCGCCCGC 2279  
Db |||||||  
4164 AGGACCCCGCAGCACCTGACCCCGCCCGCGGGGTAGCCAGCCAGCGCGCGCCCGC 4223  
QY 2280 CCTCTTCTAACCCTCAACCTTGAGCTTTCACCTGAGTCTGGCCCTCCAGCGGCGCAGG 2339  
Db |||||||  
4224 CCTCTTCTAACCCTCAACCTTGAGCTTTCACCTGAGTCTGGCCCTCCAGCGGCGCAGG 4283  
QY 2340 GAGTGGGGGATGCGGATCAGCTGGGAGGAGGAGGGGTGCTTCCACCCGAGGGGA 2399  
Db |||||||  
4284 GAGTGGGGGATGCGGATCAGCTGGGAGGAGGAGGGGTGCTTCCACCCGAGGGGA 4343  
QY 2400 AGATGCTCTCGGGACAGTTTCCCGGGCAGCTCTCGGCCAGCTTCCAGCCCGAGAGTCCCTCA 2459  
Db |||||||  
4344 AGATGCTCTCGGGACAGTTTCCCGGGCAGCTCTCGGCCAGCTTCCAGCCCGAGAGTCCCTCA 4403  
QY 2460 AGTCCAGGGCAGCTTGGGCCAGCGCAGGCGAGGAGGAGGAGTCCGAGGTGCTTGGCTTACCCCTG 2519  
Db |||||||  
4404 AGTCCAGGGCAGCTTGGGCCAGCGCAGGCGAGGAGGAGTCCGAGGTGCTTGGCTTACCCCTG 4463  
QY 2520 GGCCTCCTACTCCCCAGCACCCCTGGAGGAGGAGGGGCTCCCGCCGCGGAGGCTGCCT 2579  
Db |||||||  
4464 GGCCTCCTACTCCCCAGCACCCCTGGAGGAGGAGGGGCTCCCGCCGCGGAGGCTGCCT 4523  
QY 2580 GCCCTGGGGCCACCTCTGCATGCTGCTCATGGGGCCACCCCTGCCTCCTGGGCCCTCACTC 2639  
Db |||||||  
4524 GCCCTGGGGCCACCTCTGCATGCTGCTCATGGGGCCACCCCTGCCTCCTGGGGCCCTCACTC 4583  
QY 2640 TGCCTAGGGGAGCTGGGCCAGGCACTAGCCCTTGGCCAGGAGGTGGCCCTCAGGCTGCC 2699  
Db |||||||  
4584 TGCCTAGGGGAGCTGGGCCAGGCACTAGCCCTTGGCCAGGAGGTGGCCCTCAGGCTGCC 4643  
QY 2700 CAGGTGCCTGACCCCGAGCGGCTTCTGTGGGGCTCCCGCTCGTCAAGCTCTATCCT 2759  
Db |||||||  
4644 CAGGTGCCTGACCCCGAGCGGCTTCTGTGGGGCTCCCGCTCGTCAAGCTCTATCCT 4703  
QY 2760 GTCTGTCCCCACCCAGCTGTCCCTGCCAGGGAGCTGGGATAAAGCAGGAGGCCCGG 2819  
Db |||||||  
4704 GTCTGTCCCCACCCAGCTGTCCCTGCCAGGGAACTGGGATAAAGCAGGAGGCCCGG 4763  
QY 2820 CTCCTGGGCGAGCTGTGAGAACAGAGACTGTACCCCATCCTGCCATGCAGGCAGG 2879  
Db |||||||  
4764 CTCCTGGGCGAGCTGTGAGAACAGAGACTGTACCCCATCCTGCCATGCAGGCAGG 4823  
QY 2880 CTCTTGCCAGCCCCGTTCTGACCCCGTGTCCCGCCAGGCTGTCCCTGGGCGAAGACTCAC 2939

Db |||||||  
4824 CTCTTGCCAGCCCCGTTCTGACCCGTTTCCCCAGGCTCTGCTGGCAGAAGACTCAC 4883  
QY 2940 CTTGGAGGAGTGGGCCCTGGAGTCTGTCTCCTCCCAAGAGCCCCCAGGTTGGATTCTC 2999  
Db |||||||  
4884 CTTGGAGGAGTGGGCCCTGGAGTCTGTCTCCTCCCAAGAGCCCCCAGGTTGGATTCTC 4943  
QY 3000 AGGCTGCCAGGGCAGGCCCTCAGGAAGAAAGGGAGGCCCTTGGCTCTCCGGGAT 3059  
Db |||||||  
4944 AGGCTGCCAGGGCAGGCCCTCAGGAAGAAAGGGAGGCCCTTGGCTCTCCGGGAT 5003  
QY 3060 CAGTCTTAGGACACAGGCTCAGCCTCAGGTTGATGGGGGATGATGTGCTCCGGGGCTG 3119  
Db |||||||  
5004 CAGTCTTAGGACACAGGCTCAGCCTCAGGTTGATGGGGGATGATGTGCTCCGGGGCTG 5063  
QY 3120 CCTCTGCACGGGGCTCCACGGAGCCAGTCTCCAGACACGCTACTAAAGTGCCTAGGTT 3179  
Db |||||||  
5064 CCTCTGCACGGGGCTCCACGGAGCCAGTCTCCAGACACGCTACTAAAGTGCCTAGGTT 5123  
QY 3180 GCCGCTGTGGCTCTCCAGGGAGCAACAGAGGGCCACCAAGCAGAGGCCCTGTGGG 3239  
Db |||||||  
5124 GCCGCTGTGGCTCTCCAGGGAGCAACAGAGGGCCACCAAGCAGAGGCCCTGTGGG 5183  
QY 3240 CTGAGGATGGAGCGGCCCTCCAGCGACTCCAAGCCCGCAGAGGGCAGCCACCTGGA 3299  
Db |||||||  
5184 CTGAGGATGGAGCGGCCCTCCAGCGACTCCAAGCCCGCAGAGGGCAGCCACCTGGA 5243  
QY 3300 CTGCTCTCCTGCCAGCTGGGCTCTCTGGCTTATCTACCTTCCAGGCCACTGCAC 3359  
Db |||||||  
5244 CTGCTCTCCTGCCAGCTGGGCTCTCTGGCTTATCTACCTTCCAGGCCACTGCAC 5303  
QY 3360 TCCTGTCTGGAGGCCCTTATAGGGCAGCCAGAGCCCGCCAGCCACCCCAACAGAGA 3419  
Db |||||||  
5304 TCCTGTCTGGAGGCCCTTATAGGGCAGCCAGAGCCCGCCAGCCACCCCAACAGAGA 5363  
QY 3420 AGCACAGATCTTGGGGAGCTGCCCAACAAGCCCGCTGGCCACGAGGGCTGCAGCCGCT 3479  
Db |||||||  
5364 AGCACAGATCTTGGGGAGCTGCCCAACAAGCCCGCTGGCCACGAGGGCTGCAGCCGCT 5423  
QY 3480 GCGCTGCCGCTTCTCCCCACACCTGCCACTCCACTGTGATGTATGTCGCTCCCTC 3539  
Db |||||||  
5424 GCGCTGCCGCTTCTCCCCACACCTGCCACTCCACTGTGATGTATGTCGCTCCCTC 5483  
QY 3540 GTCTGTTCCCCAGGATCTCGAAGTGACTCCGGGCTGAGCAGTGGGGCGCTGGGGGAGG 3599  
Db |||||||  
5484 GTCTGTTCCCCAGGATCTCGAAGTGACTCCGGGCTGAGCAGTGGGGCGCTGGGGGAGG 5543  
QY 3600 GGTGACGATCTCTCAGGCTTGGCCCTGCAAGCAAAACCCACATATCTGCTCTGTATGT 3659  
Db |||||||  
5544 GGTGACGATCTCTCAGGCTTGGCCCTGCAAGCAAAACCCACATATCTGCTCTGTATGT 5603  
QY 3660 AATAAATGTCTTAACGTC 3677  
Db |||||||  
5604 AATAAATGTCTTAACGTC 5621

RESULT 5  
AB071076

LOCUS AB071076 2338 bp mRNA linear PRI 22-FEB-2003  
DEFINITION Macaca fascicularis testis cDNA clone:QtsA-16678, full insert sequence.

ACCESSION AB071076

VERSION AB071076.1 GI:15451330

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Macaca fascicularis (crab-eating macaque)

ORGANISM Macaca fascicularis

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae; Cercopitheciinae; Macaca.

REFERENCE 1

AUTHORS Osada, N., Hida, M., Kusuda, J., Tanuma, R., Hirata, M., Suto, Y., Hirai, M., Terao, K., Sugano, S. and Hashimoto, K.

TITLE Cynomolgus monkey testicular cDNAs for discovery of novel human

genes in the human genome sequence  
BMC Genomics 3 (1), 36 (2002)  
12498619  
2 (bases 1 to 2338)  
Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.  
Direct Submission  
Submitted (05-SEP-2001) Katsuyuki Hashimoto, National Institute of  
Infectious Diseases, Division of Genetic Resources; 23-1, Toyama  
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan  
(E-mail:khashi@nih.go.jp, URL:http://www.nih.go.jp/yoken/genebank/,  
Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)  
Lab host: TOP10  
Vector: pME18S-FL3 (Acc.No. AB009864)  
R. Site1: DraIII (CACTGTGTG)  
R. Site2: DraIII (CACCATGTG)  
Description: 1st strand cDNA was primed with an oligo(dT) primer  
[ATGTGCGCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized  
using specific 5' and 3' primers and amplified by PCR. The PCR  
product was digested with SfiI and size selection was performed to  
exclude fragments <1.5kb. The SfiI-digested PCR product was cloned  
into distinct DraIII sites of pME18S-FL3. XhoI sites just outside  
the DraIII sites can be used to isolate the cDNA insert. Libraries  
were constructed by oligo-capping method  
(Sugano et al., University of Tokyo, Institute of Medical Science).  
Custom primer used for sequencing  
{ 5' end primer [CTTCTGCTCTAAAGCTGCG];  
3' end primer [CGACTGCAGCTCGACACA] }.

COMMENT

FEATURES

source

1..2338  
/organism="Macaca fascicularis"  
/mol\_type="mRNA"  
/db\_xref="taxon:9541"  
/clone="QtsA-16678"  
/sex="male"  
/tissue\_type="testis"  
/clone\_lib="macaque testis cDNA library QtsA"  
/dev\_stage="adult"  
91..1890

CDS

/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="BAB64469.1"  
/db\_xref="GI:15451331"  
/translation="MKNVPVPVYCRPLVEKPTMKLWCAAGVNLGKWPSEDDGGNGV  
KPAPGRDPLTCDREGDEPKSAHGSPEKKYKELPEMDATSSRVWILTSTLTSKVVI  
IDANQPTGVDQFTVCNAHVLCISSIPAAASDYPPEGMFLDSVDNPDGVDGVLG  
ITLVGCATRCNVPRSNCSRGDTPVLDKQGEVATIANGVNPSQSTEEATEATEVPD  
PGPSEPETATLRPGPLTEHVFTDPAPTSSGPQPSGSENGPEPDSSTQPEPEPSGDPT  
GAGSSAAPTMWGAQNLWYVHSAVANWKKLHSLIKLDSVLSLVHVKGRVLVALADG  
TLAIFHRGEDQWDLNHYLMDLGHPHSIRCMVAVYDRVWCGYKHKVHVIOPKTMOI  
EKSFDHPRRESQVRLAWIGDVWVSIKLDSTLRLYHAHTHQHLDVDIEPVSKML  
GTGKLGSFVRITALLVAGSLRWGTGNGVVISIPLTETVVLHKGQLLGRANKTSPT  
SGEGARPGGIHVIYGDSDSRAASSFIYCSMAQAQLCFHGHRAVKKFFVSVPGNVLA  
TLNGSVLDSPAEPLGQLPLPRRPRVRSCTGTCW"

ORIGIN

Query Match 49.2%; Score 1816.2; DB 9; Length 2338;  
Best Local Similarity 96.2%; Pred. No. 2.5e-256;  
Matches 1871; Conservative 0; Mismatches 73; Indels 1; Gaps 1;  
QY 180 GCCTCCCTGCTCCCTGCAGTGAGTCCCAACGGGGGCCAGGAGGACACCGCGATGAAGAA 239  
Dd 39 GCCCGCCCAAGTACAAGACGCTGAGTCCCAACGGGGGCCAGGAGGACACCGCGATGAAGAA 98  
QY 240 CGTGCCCGTCCCGTGTACTGCCCGCTCTGCTGGGAGAGGACCCCAACCATGAAGCTGTG 299  
Dd 99 CGTGCCCGTCCCGTGTACTGCCCGCTCTGCTGGGAGAGGACCCCAACCATGAAGCTGTG 158  
QY 300 GTGTGCCCGGGCGTCAACTGAGCGGGTGGAGGCCCAATGAGGACGACGCTGGGAATGG 359  
Dd 159 GTGTGCTGCGGGCGTCAACTGAGCGGATGGAAGCCCACTGAGGACGACGCTGGGAATGG 218  
QY 360 AGTCAAGCCAGCGCCAGCGCCGATCCCTGACCTGCGACCGCGAAGGACGCGGAGCC 419

Db 219 AGTCAAGCCAGCGCCAGGCGCGACCCCTGACCTGGACCGCGAAGGAGACGGTGAGCC 278  
QY 420 CAAGAGCGCCACACAGCTCTCCGAGAAAGAAAGGCCCAAGAGCTCCCTGAATGACGC 479  
Dd 279 CAAGAGCGCCACACAGGCTCTCCGAGAAAGAAAGGTCAGGAGCTTCTTGAATGATGC 338  
QY 480 CACTCCAGCCGGGTGTGATCTGACCAAGCACCCTGACCAAGGAGGTCGTCATCAT 539  
Dd 339 CACTCCAGCCGGGTGTGATCTGACCAAGCACCCTGACCAAGGAGGTCGTCATCAT 398  
QY 540 CGACGCCAACACAGCCGGGACCGGTGTGGACCGATTACCGTCTGCAACCGGACGCTACT 599  
Dd 399 CGACGCCAACACAGCCGGGACCGGTGTGGACCGATTACCGTCTGCAACCGGACGCTACT 458  
QY 600 GTGATCTCCAGCATCCCGGCGCAGCAGCAGCACTACCTCCCGGGGAGATGTTCTT 659  
Dd 459 GTGATCTCCAGCATCCCGGCGCAGCAGCAGCACTACCTCCCGGGGAGATGTTCTT 518  
QY 660 GGACAGCGACGTGAACCCAGAGAGACCCGGGGCGGAGATGGCTGTCGGCCGTTATCACCT 719  
Dd 519 GGACAGCGACGTGAACCCAGAGAGACCCGGGGCGGAGATGGCTGTCGGCCGTTATCACCT 578  
QY 720 GGTGGGCTGTGCCACCCGCTGCAACGTGCGCGGAGCAACTGCTCTCCCGAGGGGACAC 779  
Dd 579 GGTGGGCTGTGCCACCCGCTGCAACGTGCGCGGAGCAACTGCTCTCCCGAGGGGACAC 638  
QY 780 CCCAGTGTAGACAAGGGGCGAGGGGAGGTGGCCACCATCGCCAAAGGGAAGTCAACCC 839  
Dd 639 CCCGCTGTAGACAAGGGGCGAGGGGAGGTGGCCACCATCGCCAAAGGGAAGTCAACCC 698  
QY 840 GTCCAGTCCACAGAGGAGGCCACAGAGGCCACGAGGTGCCAGACCTTGGGCCACGCGA 899  
Dd 699 GTCCAGTCCACAGAGGAGGCCACAGAGGCTACGAGGTGCCAGACCTTGGGCCACGCGA 758  
QY 900 GCCAGAGACGCCACATTTGCGGCCCGGCTCTTCAAGAGCAAGTCTTCACTGACCCAGC 959  
Dd 759 GCCAGAGACGCCACATTTGCGGCCCGGCTCTTCAAGAGCAAGTCTTCACTGACCCAGC 818  
QY 960 CCCGACCCCTGCTTGGCCCCCAGCTGGCAGCGGAGAACGGGCCAGAGCCTGACAGCAG 1019  
Dd 819 CCCGACCCCTGCTTGGCCCCCAGCTGGCAGTGAAGTGGGCCAGAGCCTGACAGCAG 878  
QY 1020 CAGCACACGGCCAGAGCCAGCGCCAGCGGGGACCCCAAGGAGCAGCAGTGTGCTGC 1079  
Dd 879 CAGCACACGGCCAGAGCCAGCGCCAGCGGGGACCCCAAGGAGCAGCAGTGTGCTGC 938  
QY 1080 ACCACCATGTGGTGGAGCCCAAGAACGGCTGGCTTATGTGCACTCGGCTGTGGCCAA 1139  
Dd 939 ACCACCATGTGGTGGAGCCCAAGAACGGCTGGCTTACGTGCACTCGGCTGTGGCCAA 998  
QY 1140 CTGGAAGAGTGCCTGCATCCATCAAGCTGAAGGATCTGTGCTGAGCCTGGTGCATGT 1199  
Dd 999 CTGGAAGAGTGCCTGCATCCATCAAGCTGAAGGATCTCGTGTGAGCCTGGTGCATGT 1058  
QY 1200 CAAAGCGCTGTGCTGGTGGCTCTGGCGGACGGGACCTTGGCCATCTTCCACCGTGTGA 1259  
Dd 1059 CAAAGCGCTGTGCTGGTGGCTCTGGCGGACGGGACCTTGGCCATCTTCCACCGTGTGA 1118  
QY 1260 AGATGGCCAGTGGATCTGAGCAACTATCACCTAATGACCTGGGCCACCCGCCACCTC 1319  
Dd 1119 AGATGGCCAGTGGATCTGAGCAACTATCACCTAATGACCTGGGCCACCCGCCACCTC 1178  
QY 1320 CATCCGCTGCATGGCTGTTGTGTACGACCGCGTGTGGTGGCTACAAGAACAAAGGTGCA 1379  
Dd 1179 CATCCGCTGCATGGCTGTTGTGTACGACCGCGTGTGGTGGCTACAAGAACAAAGGTGCA 1238  
QY 1380 CGTCATCCAGCCCCAAGACCATGCAGATTGAGAAAGTCAATTGACGCCCAACCCGGCGGGA 1439  
Dd 1239 CGTCATCCAGCCCCAAGACCATGCAGATTGAGAAAGTCAATTGACGCCCAACCCGGCGGGA 1298  
QY 1440 GAGCCAGGTGCGGACGTGGCGTGGATCGGCGATGGCGTATGGGTGTCCATCCCGCTTGA 1499  
Dd 1299 GAGCCAGGTGCGGACGTGGCGTGGATCGGCGATGGCGTATGGGTGTCCATCCCGCTTGA 1358

QY 1500 CTCACCCCTGAGGCTCTACCATGACACACGACACGACATCTACAGGACGTGGACATTGA 1559  
|||||  
Db 1359 CTCACCCCTGCGGCTCTACACGACACACGACACGACATCTACAGGACGTGGACATTGA 1418  
|||||  
QY 1560 GCCCTACGTCAGCAAGATGCTAGGCACTGGCAAGCTGGGTTTCTCCTTCGTACGCATCAC 1619  
|||||  
Db 1419 GCCCTACGTCAGCAAGATGCTGGGCACCGCAACTGGGTTTCTCCTTCGTGCGCATCAC 1478  
|||||  
QY 1620 GGCCTGCTTGTCCGGGCGACCGGCTCTGGGTGGGCACCGGCATCCACGTCATCTC 1679  
|||||  
Db 1479 AGCCCTGCTTGTCCGGGCGACCGGCTCTGGGTGGGCACCGGCATCCACGTCATCTC 1538  
|||||  
QY 1680 CATCCCCCTGACAGAGACTGTGTCCTGCACCGAGGCCAGCTCTCTGGGGCTCCGAGCCAA 1739  
|||||  
Db 1539 CATCCCCCTGACAGAGACTGTGTCCTGCACCGAGGCCAGCTCTCTGGGGCTCCGAGCCAA 1598  
|||||  
QY 1740 TAAGACATCCCCACCTCTGGGAGGGCGCCGCTCCCGGGGCATCATCCACGTCATGG 1799  
|||||  
Db 1599 TAAGACATCCCCACCTCTGGGAGGGCGCCGCTCTGGGGGCATCATCCACGTCATGG 1658  
|||||  
QY 1800 CGATGACAGCACTGACAGGGCGGCCAGCAGCTTCCATCCCTACTCTCCATGCGCCAGGC 1859  
|||||  
Db 1659 CGATGACAGCACTGACAGGGCGGCCAGCAGCTTCCATCCCTACTCTCCATGCGCCAGGC 1718  
|||||  
QY 1860 CCAGCTATGCTTCCATGGGCACCGCGATGCCGTAAGTTCTTTGTTCTCGTGGCAGGGAA 1919  
|||||  
Db 1719 CCAGCTCTGCTTCCACGGGCACCGCATGCTGTGAAGTTCTTTGTTCTCGTGGCAGGGAA 1778  
|||||  
QY 1920 CGTGTGGCACCCCTGAATGGCAGTGTGCTGGACAGCCCGAGCGAGGCGCCCTGGGCCAGC 1979  
|||||  
Db 1779 TGTGTGGCCACCTTGAACGGAGTGTGCTGGACAGCCCGAGTGA-GCCCTGGGCCAGC 1837  
|||||  
QY 1980 TGCCCCCTGCTCGGAGGTTCGAGGGCCAGAAAGCTGCGGAACGTGCTGTTGCTGAGCGCGG 2039  
|||||  
Db 1838 TGCCCCCTGCTCGGAGGCGCGAGGTTCAGAAAGTTCGGAACGTGCTGTTGCTGAGCGCGG 1897  
|||||  
QY 2040 GGAGGGTACATCGACTTCCGCATTGGAGACGGAGAGGACGACGAGACGGAGGCGGC 2099  
|||||  
Db 1898 GGAGGGTACATCGACTTCCGCATTGGAGACGGAGAGGACGAGACGGAGGAGGCGGCAC 1957  
|||||  
QY 2100 AGGGGACATGAGCCAGGTGAAGCCC 2124  
|||||  
Db 1958 AGGGGACATAAGCCAGGTGAAGCCC 1982  
|||||

RESULT 6  
BC060603  
LOCUS BC060603 5440 bp mRNA linear ROD 12-NOV-2003  
DEFINITION Mus musculus mitogen-activated protein kinase 8 interacting protein  
3, mRNA (cDNA clone MGC:79152 IMAGE:5706805), complete cds.  
ACCESSION BC060603  
VERSION BC060603.1 GI:38148641  
KEYWORDS MGC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 5440)  
Straussberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,  
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,  
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,  
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,  
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,  
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,  
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,  
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,  
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,  
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,  
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S.,  
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,

Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,  
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,  
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,  
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
22388257  
12477932  
2 (bases 1 to 5440)  
Straussberg,R.  
Direct Submission  
Submitted (31-OCT-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: http://mgc.nci.nih.gov  
Contact: MGC help desk  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
cDNA Library Preparation: M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (L1NL)  
DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.  
Thomas L. Casavant.  
Web site: http://genome.uiowa.edu  
Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu  
Bonaldo,M.F., Akabogu,I., Bair,T., Bair,J., Crouch,K., Davis,A.,  
Fishler,K., Keppel,C., Kucaba,T., Lebeck,M., Melo,A., Schaefer,K.,  
Scheetz,T., Smith,C., Snir,E., Tack,D., Trout,K., Walters,J.,  
Casavant,T., Soares,M.B.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/L1NL at: http://image.lnl.gov  
Series: Plate: Row: Column: 0  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 7305254.

FEATURES  
source  
1. .5440  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CS7BL/6"  
/db\_xref="taxon:10090"  
/clone="MGC:79152 IMAGE:5706805"  
/tissue\_type="Brain, mouse 15.5 dpc"  
/clone\_lib="NIH\_BMAP\_EX0"  
/lab\_host="DH10B"  
/note="Vector: pYX-ASC"  
1. .5440  
/gene="Mapk8ip3"  
/note="synonyms: Jip3, JSAP1, JSAP1a, JSAP1b, JSAP1c,  
JSAP1d, Syd2, mKIA1066"  
/db\_xref="LocusID:30957"  
/db\_xref="MGI:1353598"  
59. .4003  
/codon\_start=1  
/product="Mapk8ip3 protein"  
/protein\_id="AAH60603.1"  
/db\_xref="GI:38148642"  
/db\_xref="LocusID:30957"  
/translation="MMEIQMDEGGVVVYQDDYCSGVMSESVGLAGSIYREFERLI  
HCYDEVVVELMPLVNVLENLDSVLENEHEVELELRENEQLLTQYEREKALRK  
QAEKEFTFEDALEQEKELQIQVEHYEFQTRQLELAKYADQISRLEERESEMKKE  
YNALHQHTEMIQTYVHERSKMQVGSGQTSSLPGRSRQSWRKSRKERPTSLN  
VFPLADGCPNDEMSESGSSAAATPTTGTSTNTPTSSVPSAAVTPLESLOPLGDI  
VSVTKNNQAREKRNMEVQVTQEMRNVSIGMSSDEWSDVDQIIDSTPELDVCP  
TRLERTSSPTQIVNKAFGINTDSLYHELSTAGSEVIGDVDEGADLLGEFVSRDDFF  
GMGKEVGNLLENLQLETKNALNVKNDLIAKVDLQSGEVLKGELEAAKQAKVL  
ENRIKELEELKRVKSEAVTARPREVEDEVSSYLCTELDKIPMAQRRFTRVEMAR  
VLERNQYKERLMELQEAVRTWEMIRASREHPSVQEKKSTIWQFFSRLFSSSSPP  
AKRSYPSVNIHYKSPTAAGFSQRRSHALCQISAGSRPLEFFPDCTSSARREQREQ  
YRQREVRNDDGRLQACGWSLPAKYQLSPNGGDETRMKNVPVVPVCRPLVEKDPS  
TKLWCAAGVNLGKWKHEEDSSNGPKPVGDRDPLTCDREGEPEKSTHPSPEKKAKE  
TPEADATSSRWILTSTLTTSKVIIIDANPGTIVDQFTVCNAHVLICISSIPAASDSD

CDS

gene

YPPCEMFLDSDVNPEDSGADGVLAGITLVGCATRONVPRNSCSRGRDTPVLDDKGGQDV  
ATTANGKVNPSQSTEEATEATEVPDPGPSESEATVRPGPLTEHVFTDPAPTPSSSTQ  
PASENGESNGTIQOPQVEPSGELSTTSAAAPTWMWGAQNGWLYVHSASAVANWKCLH  
SIKLDSVLSLVHVKGRLVALADGTLAIFHRGEDQWDLSNYHLMDLGHPHHSIRCM  
AVNDRVWCYGNKVHVIOQKTMQIEKSFDAHPRPRESQVROLAWIGDGVWVSIRLDST  
LRLYHAHTHOHLQDVIDIEPYVSKMLGTGKLGFSFVRITALLIAGNRLWVGTGNGVVIS  
IPLTETVWLHRGQLGLRANKTSPTSSEGTSPGGIHHVYGDSDSKAASFIPIYCSMA  
QALCFHGHRDAVKFFVSVPGNVLATLNGSVLDSFSEGPGPAAPRAADAEGQKLNALV  
LSGGBGYIDFRIGDGEDDETEECAGDVNQTKPSLSKAERSHIIIVWQVSYTPE"  
254. .598  
/note="filament; Region: Intermediate filament protein"  
/db\_xref="CDD:pfam00038"  
1220. .1672  
/note="filament; Region: Intermediate filament protein"  
/db\_xref="CDD:pfam00038"

ORIGIN

|                       |              |   |  |                      |
|-----------------------|--------------|---|--|----------------------|
| Query Match           | 48.9%;       | Score 1807;   | DB 10;   | Length 5440;         |
| Best Local Similarity | 73.1%;       | Pred. No. 4.2e-255;   |  |                      |
| Matches 2605;         | Conservative | 0;  | Mismatches 850;                                | Indels 108; Gaps 18; |
| Qy                    | 148          | GATGGGTAGAGCCAGGGTTCGTGCCCCACGGCGCCTCCCTGCTCCCTGCAGCTGAGTCCC    | 207  |                      |
| Db                    | 1964         | GACGGAGGCTGCAGGCCTGTGGGTGGAGCCTGCCTGCCAAGTACAAGCAGCTGAGCCCC     | 2023   |                      |
| Qy                    | 208          | AACGGGGGCCAGGAGGACACGGGATGAAGAACGTGCCGGTGCCGGTGTA               | 267  |                      |
| Db                    | 2024         | AATGGAGGCCAGGAAGACACCCGGATGAAAAATGTGCCTGTCCCTGTGTACTGCCGCCCT    | 2083   |                      |
| Qy                    | 268          | CTGGTGGAGAAAGACCCCAACCATGAAGCTGTGGTGTGCCGCGGGCGTCAACCTGAGCGGG   | 327  |                      |
| Db                    | 2084         | CTGGTGGAGAAAGACCCCTTCGACAAAGCTGTGGTGTGCTGTGGTGTCAACCTGAGTGGG    | 2143   |                      |
| Qy                    | 328          | TGAGAGCCCAATGAGGACGACGCTGGGAATGGAGTCAAGCCAGCGCCAGCGCGGATCCC     | 387  |                      |
| Db                    | 2144         | TGGAAGCCCATGAAGAGGA   | CTTAGCAATGGACCAAGCCTGTACCAAGGTCGAGGTCGAGACCCCT | 2203                 |
| Qy                    | 388          | CTGACCTGCGACCGCGAAGGAGACGGCGAGCCCAAGAGCGGCCACACAGCTCTCCCAGAGAAG | 447  |                      |
| Db                    | 2204         | CTGACCTGTGACCGGGAAGGAGAGGCGGAACCCAAAGAGCACACACCCATCACCTGAGAAG   | 2263   |                      |
| Qy                    | 448          | AAGAAGGCCAAGGAGCTCCCTGAAATGGACGCCACCTCCAGCGCGGTGTGGATCCCTGACC   | 507  |                      |
| Db                    | 2264         | AAGAAGGCCAAGGAAACCCCTGAGGCAGATGCTACTCCAGTCCGGTATGGATCCCTCACC    | 2323   |                      |
| Qy                    | 508          | AGACCCCTGACACACAGCAAGGTGTGTATCATCGACGCCCAACAGCGCGGCACGGTGGTG    | 567  |                      |
| Db                    | 2324         | AGCACCTTGACAAACAGCAAGGTGTGTATCATTTGATGCCAACAGCCAGGCACAAATTGTG   | 2383   |                      |
| Qy                    | 568          | GACAGATTACCGTGTGCAACGGGACCGTGTGTGATCTCCAGATCCCCCGGGCCAGC        | 627  |                      |
| Db                    | 2384         | GATCAGTTCACAGTCTGCAATGCCACGTCCTGTGTATCTCCAGCATTCCTGCGGCCAGT     | 2443   |                      |
| Qy                    | 628          | GACAGCGACTACCTCCCGGGGAGATGTTCTCTGGACAGCGACGTGAACCCAGAGGACCCG    | 687  |                      |
| Db                    | 2444         | GACAGTGACTATCCCTCTGGGAGATGTTCTCTAGACAGTGTGTGAACCCCTGAAGATTCA    | 2503   |                      |
| Qy                    | 688          | GGCGCAGATGGCGTGTGGCCGGTATCACCTGTGTGGCTGTGCCACCCGCTGCAACGTG      | 747  |                      |
| Db                    | 2504         | GGTGTGATGGTGTGTGGCTGGCATCACCTGTGTGGGTGTGTACTCCGCTGCAATGTT       | 2563   |                      |
| Qy                    | 748          | CCGCGAGCAACTGCTCTCCCGAGGGGACACCCCAAGTCTAGACAGAGGGGAG            | 807  |                      |
| Db                    | 2564         | CCACGTAGCAACTGTTCTCACGAGGAGACACCCCAAGTACTGGACAAGGGGCAGGGGGAT    | 2623   |                      |
| Qy                    | 808          | GTGGCCACCATCGCCAAAGGGGAGGTCAACCCGTCCTCCAGTCCACAGAGAGGCCACAGAG   | 867  |                      |
| Db                    | 2624         | GTGGCCACCACTGCCAATGGGAAGTCAACCCGTCCCAATCCACAGAGAAGCCACAGAA      | 2683   |                      |
| Qy                    | 868          | GCCACGAGGTGCCAGACCTTGGGCCAGCGAGCCAGACAGACCCACATTGCGGCCCGGG      | 927  |                      |
| Db                    | 2684         | GCCACGAGGTGCCAGACCTTGGTCCACGCGAGTCAAGAGCAACGACAGTCCGGCCCGGG     | 2743   |                      |

|    |      |  |      |  |
|----|------|--|------|--|
| Qy | 928  | CCTCTCACAGACACGTCTTCACTGACCCAGCCCCGACCCCGTCCCTCTGGCCCCCAGCCT   | 987  |  |
| Db | 2744 | CCTCTCACAGACATGTCTTTACTGACCCAGACCCACCCCATCTCTCCAGCACCCAGCCT    | 2803 |  |
| Qy | 988  | GGCAGCGAGAACGGGCCAGACGCTGACAGCAGCAGCACACGGCCAGAGCCAGAGCCCCAGC  | 1047 |  |
| Db | 2804 | GCCAGTGAGAAATGGGTCAAGTCCAATGGCAACATTGTACAGCCTCAGGTGGAGCCCACT   | 2863 |  |
| Qy | 1048 | GGGACCCACCGGAGCAGGAGCAGTGTCTGACCCACCCATCATGTGGCTGGGAGCCCAAG    | 1107 |  |
| Db | 2864 | GGGGAACCTCTCAACAACAACAGTAGCGCTGCACCCACTATGTGGCTAGGAGCCCAAT     | 2923 |  |
| Qy | 1108 | GGCTGGCTCTATGTGCACCTCGGCTGTGGCCAACTGGAAGAAGTGCCTGCATCAATCAAG   | 1167 |  |
| Db | 2924 | GGCTGGCTCTATGTGCATTCAGCGGTAGCCAACTGGAAGAAGTGTCTGCACCTCCATCAAG  | 2983 |  |
| Qy | 1168 | CTGAAGGATTCTGTGCTGAGCCTGGTGCATGTCAAAAGCCGCTGTGCTGGTGGCTCTGGCG  | 1227 |  |
| Db | 2984 | CTAAAAGACTCTGTGCTGAGCCTGGTGCATGTCAAAGGCCGAGTGTCTGGTAGCTCTTGCA  | 3043 |  |
| Qy | 1228 | GACGGGACCTGGCCATCTTCCACCGTGGTGRAGATGGCCAGTGGGATCTGAGCAACTAT    | 1287 |  |
| Db | 3044 | GATGGGACCTTGGCTATCTTCCATCGTGGAGAGGATGGCCAGTGGGACCTGAGCAACTAC   | 3103 |  |
| Qy | 1288 | CACCTAATGGACCTGGGCCACCCGACCACTCCATCCGCTGCATGGCTGTTGTGTAGCAC    | 1347 |  |
| Db | 3104 | CACCTAATGGACCTGGGCCACCCACACCACTCCATCCGCTGCATGGCTGTTGTGAATGAC   | 3163 |  |
| Qy | 1348 | CGCGTGTGGTGTGGCTACAAGAAACAGGTGCACGTCTATCCAGCCCCAAGACCATGCAGATA | 1407 |  |
| Db | 3164 | CGAGTTTGGTGTGGCTACAAGAAACAGGTGCATGTTATCCAGCCCCAAGACAATGCAGATT  | 3223 |  |
| Qy | 1408 | GAGAAATCATTTGATGCCCAACCCGCGGGGAGAGCCAGGTGCGGAGCTGGCGTGGATC     | 1467 |  |
| Db | 3224 | GAGAAATCATTTGATGCCCAACCCGCGGGGAGAGCCAGGTACGTACGTGGCTGGATC      | 3283 |  |
| Qy | 1468 | GGCATGGCGTATGGGTGTCCATCCGCGCTGGATCCACCCCTGAGGCTCTACCATGCACAC   | 1527 |  |
| Db | 3284 | GGTATGGAGTGTGGGTCTCTATTGCTTGGATTCTACCCCTCGGCTCTACCATGCTCAC     | 3343 |  |
| Qy | 1528 | ACGCACAGCATCTACAGGACGTGGACATTGAGCCCTACGTAGCAAGATGCTAGGCAC      | 1587 |  |
| Db | 3344 | ACCCACAGCACCTGCAGGATGTGGACATTGAGCCCTATGTTAGCAAGATGCTAGGAACC    | 3403 |  |
| Qy | 1588 | GGCAAGCTGGTTCCTTCTGTAAGCATACCGGCCCTGCTTGTGCGGGCAGCCGGCTC       | 1647 |  |
| Db | 3404 | GGCAAGCTGGGCTTCTCTTCTGTCGCATCACAGGCTTACTCATTCAGGCAACCGTCTG     | 3463 |  |
| Qy | 1648 | TGGTGGGCACCCGGCAACGGAGTGGTCTATCTCCATCCCCCTGACAGAGACTGTGGTCTG   | 1707 |  |
| Db | 3464 | TGGTGGGCACCTGGCAATGGGGTTGTCTATCTCCATCCCCCTTGA                  | 3523 |  |
| Qy | 1708 | CACCGAGGCCAGCTCTGGGCTCCGAGCCCAATAGACATCCCCACCTCTGGGGAGGGC      | 1767 |  |
| Db | 3524 | CATCGAGGCCAGCTCTAGGGCTCCGAGCCCAACAAGACATCCCCAACA               | 3583 |  |
| Qy | 1768 | GCCCTCCCGGGGCATCATCCACGTGTATGGCGATGACAGAGTGACAGGGCGGCCAGC      | 1827 |  |
| Db | 3584 | ACCCGCCAGGGGCATCATCCATGTGTATGGGACGACAGCAGTGACAAGGCCGCCAGT      | 3643 |  |
| Qy | 1828 | AGCTTCATCCCTACTGCTCCATGGCCAGGCCAGTATGCTTCCATGGGCACCCCGAT       | 1887 |  |
| Db | 3644 | AGTTTCATCCCTACTGCTCCATGGCACAGGCTCAGCTTGTCTTCCATGGGCACCCGAT     | 3703 |  |
| Qy | 1888 | GCCGTGAAGTTCTTTGTCTCGGTGCCAGGGAACGTGTGGCCACCCCTGAATGGCAGTGTG   | 1947 |  |
| Db | 3704 | GCTGTCAAAATTTCTTGTCTCTGTGCCAGGAAATGTGTGGCCACTCTCAATGGCAGTGTG   | 3763 |  |
| Qy | 1948 | CTGGACAGCCCCAGCCGAGGGCCCTGGGCCAGCTGCCCTCGGAGTCCGAGGTCGAGGGCCAG | 2007 |  |
| Db | 3764 | CTAGACAGCCCCATCAGAGGGCCCTGGGCCCTGTGCACCCGCTGCAGATGCTGAGGGCCAG  | 3823 |  |
| Qy | 2008 | AAGCTCGGGAACGTGTGTGTGCTGAGCGCGGGGAGGGSCTACATCGACTTCCGCAATTGGA  | 2067 |  |



|      |  |  |
|------|--|--|
| gene | 1. .5443   |  |
|      | /gene="Jip3"   |  |
| CDS  | 41. .4009  |  |
|      | /gene="Jip3"   |  |
|      | /function="JNK MAP kinase signal transduction pathway  |  |
|      | molecular scaffold"  |  |
|      | /note="JIP3a; putatively alternatively spliced"  |  |
|      | /codon_start=1   |  |
|      | /product="JNK interacting protein-3a"  |  |
|      | /protein_id="AAF26842.1"   |  |
|      | /db_xref="GI:6724092"  |  |
|      | /translation="MMELIQDEGGVVVYQDDYDCSGVMSERVSGLAGSIYREFERLI<br>HCYDEVVKELMPLVNVNLENLSDVLSNQEHEVELELLREDNEQLLTQYEREKALRK<br>QAEKFI EFEDALEQEKELQIQVEHYEFQTRQLELKAKNYADQISRLERESEMKKE<br>YNALHORHTEMIQTVVEHIERSKMQVGGSGQTSSLPGRSRKERPTSLNVFPLADGM<br>VRAQMGGKLVPA GDHWHLSDLGLQSSSSYQCPNDEMSESGOSSAATPSTGTKSNIT<br>PTSSVPSAAVTPLNESIQPLGDYVSVTKNNKQAREKNSRNMEVQVTQEMRNVSIGMG<br>SSDEWSDVDIIDSPELDVCPETRLERTGSSPTQGVNKAFGINTDSLHELSTAGS<br>EVIQDVEGADLLGFESGMGKEVGNLLLENSQLLETKNALNVVKNDLIAKVDQLSGEQ<br>EVLKGLEAAKQAKVLENRIKELEELKRVKSEAVTARPREVEEDDKIPMAQRRR<br>FTRVENARVLMERNQYKERLMELQEAVRWTEMIRASREHPSVQEKKXSTIWQFFSRLF<br>SSSSPPPAKRSPSVNIHYKSPTAAGFSQRRSHALQIISAGSRPLEFFPDDDDCTSSA<br>RREQKREQYRQVREHVRNDDGRLOACGWSLPKAKYKQLSPNGGQEDTRMKNVPVPVYCR<br>PLVEKDPSTKLWCAAGVNLSGWKPHEDSSNGPKVPGRDPLTCDREGEGEPKSTHPS<br>PEKKAKETPEADATSSRVWILTSTLTTSKVIIIDANQPGTIVDQFTVCNAHVLCISS<br>IPAAASDYPGEMFLDSVNPEDSGADVLGITLVGCATRCNVPRNSCSSRGDTPV<br>LDKGQGVATTANGKVNPSQTEATEATEVPDPGSPSEATTVRPGPLTEHVFTDPA<br>PTPSSSTQPASENGSENGTIVQPOVEPSGELSTTTSSAAPTMWLGAQNGWLYVHSVA<br>ANWKKCLHSIKLSDSVLSLVHVKGRVLVALADGTLAIFHRGEDQWDLNSYHLMDLGH<br>PHHSIRCMVAVNDRVWCYKNKVHVIQPTMQIOIEKSFDAHPRESQVQLAWIGDGVW<br>VSIRLDSILRLYHAHTHOHQVDVIEPYVSKMLGTGKLFSGFVRITALLIAGNRLWVG<br>TNGVVISIPLTETVHLRGQLLGRANKTSPTSSEGTTPGGIIHVYGDSSDKAASS<br>FIPYCSMAQAQLCFHGHRAVKFFVSPGNVLATLNGSVLDSPESEGPAPAAPDAEG<br>QKLKNALVLSGGEGYIDFRIGDGEDDETEECAGDVNQTKPSLSKAERSHIIIVQVSYT<br>PE" |  |
|      | ORIGIN   |  |
|      | Query Match  | 48.7%; Score 1797.4; DB 10; Length 5443;                                 |
|      | Best Local Similarity  | 72.9%; Pred. No. 1.1e-253;   |
|      | Matches 2599; Conservative   | 0; Mismatches 856; Indels 108; Gaps 18;                                  |
|      | Qy   | 148 GATGGGTAGGAGCCAGGGTTCGTGCCACAGGGCCCTCCCTGCTCCCTGCAGCTGAGTCCC 207<br> |

|    |      |   |
|----|------|---|
| Qy | 628  | GACAGGACTACCTCCCGGGAGATGTTCTCTGGACAGCGACGTGAACCCAGAGGACCCG 687        |
| Db | 2450 | GACAGTGACTATCCCCCTGGGGAGATGTTCTCTAGACAGTGATGTGAACCTTGAAGATTCA 2509    |
| Qy | 688  | GGCGCAGATGGCGTGCTGGCCGGTATCACCCCTGGTGGCTGTGCCACCCCGCTGCAACGTG 747     |
| Db | 2510 | GGTGTGATGGTGTGCTGGCTGGCATCACCCCTGGTGGGTGTGCTACCCCGCTGCAATGTT 2569     |
| Qy | 748  | CCGCGGAGCAACTGCTCCTCCGAGGGGACACCCCCAGTGTAGACAAAGGGGACGGGGAG 807       |
| Db | 2570 | CCACGTAGCAACTGTTCTCTACGAGGAGACACCCCACTACTTGAACAAGGGCAGGGGGAT 2629     |
| Qy | 808  | GTGGCCACCATCGCCAAACGGGAAGGTCAACCCCGTCCCAGTCCACAGAGGAGGCCACAGAG 867    |
| Db | 2630 | GTGGCGACCACTGCCAATGGGAAGGTCAACCCCGTCCCACATCCACAGAAGAAGCCACAGAA 2689   |
| Qy | 868  | GCCACGGAGGTGCCAGACCCCTGGGCCCCAGCGAGCCAGAGACAGCCACATTTCGCGCCCCGG 927   |
| Db | 2690 | GCCACAGAGGTGCCAGACCCCTGGTCCCAGCGAGTCAGAAGCAACGACAGTCCCGSCCGGG 2749    |
| Qy | 928  | CCTCTCACAGACGACGTCTTCACTGACCCAGCCCCGACCCCGTCTCTTGGCCCCCAGCCT 987      |
| Db | 2750 | CCTCTCACAGAGCATGTCTTTACTGACCCAGCACCCCACTCTCCAGCACCCAGCCT 2809         |
| Qy | 988  | GGCAGCGAGAACCGGGCCAGAGCTGTACAGCAGCAGCAGCACACGGCCAGAGCCAGAGCCCAGC 1047 |
| Db | 2810 | GCAGTGAGAAATGGGTGAGATCCAATGGCAACCATTTGTACAGCCTCAGGTGGAGCCCAGT 2869    |
| Qy | 1048 | GGGACCCCCACGGGAGCAGGACGAGTGTGCTGACACCCACCATGTGGTGGGAGCCCAGAAC 1107    |
| Db | 2870 | GGGAAACTCTCAACAACAACAGTAGCGCTGCACCCACTATGTGGTAGGAGCCCAGAAAT 2929      |
| Qy | 1108 | GGCTGGCTCTATGTGCACTCGGCTGTGGCCAACTGGGAAGAAGTGCCTGCACCTCCATCAAG 1167   |
| Db | 2930 | GGCTGGCTCTATGTGCATTGACGGGTAGCCAACTGGGAAGAAGTGTGTGCATCCATCAAG 2989     |
| Qy | 1168 | CTGAAGGATTCTGTGCTGAGCCTGTGTCATGTCTAAAGCCCGTGTGTGGTGGTCTGTGGCG 1227    |
| Db | 2990 | CTAAAGACTCTGTGCTGAGCCTGTGTCATGTCTAAAGCCCGAGTGTGTGGTAGCTCTTGCA 3049    |
| Qy | 1228 | GACGGACCCCTGGCCATCTTCCACGTGGTGAAGATGGCCAGTGGGATCTGAGCAACTAT 1287      |
| Db | 3050 | GATGGACCCCTGGCTATCTTCCATCTGTGGAGAGATGGCCAGTGGGACCTGAGCAACTAC 3109     |
| Qy | 1288 | CACCTAATGGACCTGGGCCACCCCGACCACTCCATCCGCTGCATGGCTGTGTGTGACGAC 1347     |
| Db | 3110 | CACCTAATGGACCTGGGCCACCCCACTCCACTCCATCCGCTGCATGGCTGTGTGTGATGAC 3169    |
| Qy | 1348 | CGCTGTGGTGTGGCTACAAGAACAGGTGCACGTCTATCCAGCCCAAGACCATGCAGATA 1407      |
| Db | 3170 | CGAGTTGGTGTGGCTACAAGAACAGGTGCATGTTATCCAGCCCAAGACAATGCAGATT 3229       |
| Qy | 1408 | GAGAAATCATTTGACGCCCAACCGCGCGGGAGAGCCAGGTGCGGAGCTGGCGTGGATC 1467       |
| Db | 3230 | GAGAAATCATTTGATGCCCAACCAAGCGGGGAAAGCCAGGTACGTGAGTGGCTGGATC 3289       |
| Qy | 1468 | GGCATGGCGTATGGGTGTCCATCCGCTGGACTCCACCCCTGAGGGCTTACCATGCACAC 1527      |
| Db | 3290 | GGTGTGGAGTGTGGGTCTCTATTCCGCTTGGATTCTACCCCTTGGCTTACCATGCTCAC 3349      |
| Qy | 1528 | ACGCACAGCATCTACAGGACGTGGACATTTAGCCCTAGCTCAGCAAGATGCTAGGCACT 1587      |
| Db | 3350 | ACCCACAGCACCTGCAGGATGTGGACATTTAGCCCTTACTCTATTGCAAGATGCTAGGAACC 3409   |
| Qy | 1588 | GGCAAGTGGGTTTCTCCTTGTACGCATCACGGCCCTGCTTGTGCGGGGAGCCCGGCTC 1647       |
| Db | 3410 | GGCAAGTGGGCTTCTCCTTGTGCGCATCACAGCCTTACTCTATTGCAAGATGCTAGGAACC 3469    |
| Qy | 1648 | TGGGTGGGACCGGCAACGGAGTGGTCTATCTCCATCCCCCTGACAGAGACTGTGGTCTTG 1707     |
| Db | 3470 | TGGGTGGGCACTGGCAATGGGGTTGTCTATCTCCATCCCTTGAAGTGTGACTGTGGTCTTG 3529    |
| Qy | 1708 | CACCGAGGCCAGCTCCTGGGGCTCCGAGGCCAATAAGACATCCCCCACCTCTTGGGGAGGGC 1767   |

Db 3530 CATCGAGGCCAGCTCCTAGGGCTCGAGCCAAACAGACATCCCCAACATCTGGGGAGGGG 3589  
QY 1768 GCCCGTCCCGGGGCATCATCCAGTGTATGGCGATGACAGCAGTGACAGGGCGGCCAGC 1827  
Db 3590 ACCCGCCAGGGGCATCATCCATGTGTATGGGACGACAGCAGTGACAAAGGCCGCCAGT 3649  
QY 1828 AGCTTCATCCCTACTGCTCCATGGCCAGGCCAGCTATGCTTCCATGGGCACCGCGAT 1887  
Db 3650 AGTTTCATCCCTACTGCTCCATGGCACAGGCTCAGCTTTGCTTCCATGGGCACCGTGAT 3709  
QY 1888 GCCGTGAAGTTCTTGTCTCGGTGCGCAGGAAACGTGCTGGCCACCTGAAATGGCAGTGTG 1947  
Db 3710 GCTGTCAAATTTCTTGTCTCTGTGCCAGGAAATGTGCTGGCCACTCTCAATGGCAGTGTG 3769  
QY 1948 CTGGACAGCCAGCCGAGGGCCCTGGGCCAGCTGCCCTGCCCTCGGAGGTCGAGGGCCAG 2007  
Db 3770 CTAGACAGCCCATCAGAGGGCCCTGGGCCCTGCTGCACCCGCTGCAGATGCTGAGGGCCAG 3829  
QY 2008 AAGCTGCGGAACGTGCTGGTGTGAGCGCGGGAGGGCTACATCGACTTCCGCATTGGA 2067  
Db 3830 AAGTTGAAGAAATGCACTGGTGTGAGTGGTGTGAAGTTTACATTGACTTCCGTATCGGA 3889  
QY 2068 GACGGAGAGGACGACGAGACGAGGAGGGCGCAGGGGACATGAGCCAGGTGAAGCCCGTG 2127  
Db 3890 GACGGAGAGGATGATGAACCTGAGGAATGTGCCGGGACGTGAACACGACAAAGCCCTCG 3949  
QY 2128 CTGTCCAAGGACAGCGCAGTCACATCATCGTGTGGCAGTGTCTTACACCCCGAGTGA 2187  
Db 3950 TTGTCCAAGGCTGAGCGCAGCCACATCATCGTGTGGCAGTGTCTTACACCCCTGAGTG- 4008  
QY 2188 AGCTGTGCTGCTGCGCCGACCTGTACATPAGGACCCCGACCACTGACCCCGCCG 2247  
Db 4009 AGACCTGTCTACCTGATGATGCAACTGTACATAGGACCCCTACCTGCCCTGCCCGCT 4068  
QY 2248 GGGCCGCGGGGTAGCCAGCCAGCGCCCGCCGCTCTTCTAAACCTCTCAACCTGCAGCT 2307  
Db 4069 GTTCCCTGGGGCAGCCA-----GGTTCGTCCATCCCTTTTAACTCTCACTTGCAGCT 4123  
QY 2308 TTCACCTGAGTGTGCCCCCTCCAGCGGGCAGGAGTGGGGATGCGGATCAGCTGGGAG 2367  
Db 4124 TTTGCTTGAAGTCCAGCCCTAGCTGTTAGAGAGGATGAAGCAGGGTTTGGGGAGTGAG 4183  
QY 2368 GAGGAGGGAGGGGTGCTTCCACCCGAGGGGAAGATGCTCTCGGGACAGTTTCCCGGCA 2427  
Db 4184 ACAGGAACAGATGTCTTTCCCATCAGGAGGAGAGAGCCCTGTGAGACATCTCTCCAGCA 4243  
QY 2428 GCTCCTGGCCAGCTTCCAGCCCAAGT- CCTCAAGTCCAGGGCACCTTGGGCCAGCGCA 2486  
Db 4244 GCGCCCTGCTTTGACCCCAACATGAAGTCCAGGGGGCCCTCAGGGCAAGCAGGACCAAA 4303  
QY 2487 GGCAGATCCGAGGTGCTGCTGCTACCTTGGCCCTCTACTCTCCCGACGACCCCTGGA 2546  
Db 4304 ATGACTATGGGCAGGCCCTTACAGGTGGCCCTGGCTCCACTCCATCTCACCAGTCTTGGC 4363  
QY 2547 GGAGGAGGGGCTCCCGCGCCGAGGCTGCTGCTGCGCCCTGGGCCCACTCTGCTATGCTG- 2604  
Db 4364 AGGAGCAAGAGCCTCATACTGCCAGAGTCTTCCCAAGTCTTCCCAAGTCTACTTGCCTTGCATG 4423  
QY 2605 CTCATGGGGCCACCTTCTCTGGGCTTCACTCTGCTAGGGAGTGGGGAGTGGGCCAGGCAC 2664  
Db 4424 CTGAGGAAGGGCTCCTGATTTCCAAACCTTTGCTCTCTCTGAGGGGCCAGGGCTTAGCCA 4483  
QY 2665 TAGCCTTTGCCCCAGGGAGGTGGGCTCAGGCTGCCAGGTGCCCTGCAACCCAGCCGGCCT 2724  
Db 4484 CCGTCTCTGCCCCAGAGCAAGGTTAGGGCTACCTGTGTTCCCGGGGCCCCAGGACGCT 4543  
QY 2725 TCTCTGGGGCCCTCCCGCTGCTCAAGCCTCTATCTGTGTGTGCCCACTCTGCTGCTGCTG 2784  
Db 4544 TCTCT- GGGCTTCCCGCTCTCCAACTTATATTCCTTGGCCCTTAGCCTGGGTCTTACC 4602  
QY 2785 TGCCCCAGGAGCTGGCATAAAGCACAGGCCCGGCTCCCTGGGGCAGCTGCTTGAGAAC 2844

Db 4603 TG-----GGAACCTGGCATGAAAGCTCAGGGCCAGCCTCCCAAGGSCAGCTGCTTCAAAG 4658  
QY 2845 AGAGACTGCTACCCCATCTGCCCATGCGCATGCGGAGGCTCTTTGCCAGCCCCCTTCTGACCCG 2904  
Db 4659 A-----TGACAAATACCCGTCAACCCAAAGATTGTCAGCCCCCTCACAGACCAG 4706  
QY 2905 TGTCCCCCAGGCTCTGCTGGCAGAAAGACTCACCTTTGGAGGAGTGGGCCCTGAGAGTCC 2964  
Db 4707 TTCCCTCTCC--GCTCTTCTGGACACAAGATTACCTTTGGAGGAG-----CAAGTGCCC 4758  
QY 2965 TGTCCCTCCAGAACCCCCAGGCTGGGATTCTCAGGCTGCCAGGGCAGGCCAGGCT 3024  
Db 4759 TGCCCTCTCCAGAAAGCCCCAAGGTAGAATTTCTCAGGCTGCCAGGGCAGGCACAGGCT 4818  
QY 3025 CAGGAAGAAGGGAGGCCCTTGGCCTCTCCGGGATCAGTCTTAGGACACAGGCTCAGCCT 3084  
Db 4819 CAGGAAGAAGGG--TGGCCCTGGTCTCCCCCAGAGTCACTCTTAGGATGCAGGCTTAGCCT 4876  
QY 3085 CAGGTTGATGGGGATGATGTCTCCCGGGCTGCCCTCCTGCACGGGGCTCCACGGAGC 3144  
Db 4877 CAGGTTAATGGAGGATGTTGTCTCCAAGGCTGCCCTCCTGCACAGGGCTCCAAGGAGC 4936  
QY 3145 CCAGCTCCAGACACGCTACTAAGTGCCTAGGTTGCCCGCTGTGGCTGTGCCCTGCTCCAGGG- 3203  
Db 4937 CCAGCTCCAGACACG-TACTAAGTGCCTAGGTTGCCGGTGTGGCTGTGCCCTGCTCCCTTGA 4995  
QY 3204 ----AGCAACAGAGAGGCCCAAGCAGAGGCCCGTGGGGCTGAGGATGGAGCGGCCCCC 3259  
Db 4996 GAGCAGCTGACAAAGAGGCTTCCAGGCAGAGGCCAGAGAGCTGAGGA-----CCAGCC 5048  
QY 3260 AGCCGACTCCAAAGCCCGCAGAGGGCAGACGCCACCTGGACTGCTCTCCCTGCCCCAGCTG 3319  
Db 5049 AGGATGGGCACAGTTCTAAGCAAGTGTGTGGAGCAGGTGTGCCCTGGCTGTGGCTTGG 5108  
QY 3320 GGCCTCTCTGGCCCTATTCTTACCTTCCAGGCCCACTGCACTCTCTGTCTGGGAGGCCCTTA 3379  
Db 5109 CTCGTACGGGC-----A 5121  
QY 3380 TGAGGGCAGCCCCAGCCCCCGCACCCCAACCCCAACAGAGAACACAGATCTTGGGGAGCT 3439  
Db 5122 CTGAGTCAAGTTGGCCTCTACTGCAAGCTTCCACCAAGAGACACAGATCCAGGGGCACT 5181  
QY 3440 GCCCCACAGACCCCGCTGGCCACCGAGGCTGCAGC--CGCTGCGTGGCGGCTTCTCCC 3497  
Db 5182 GCCCC-TGAGCCTGTCTGGCCACAAACAGCTGCAGAGTGTGCAGCTGCAGGCTTCTCCC 5240  
QY 3498 CACCACCTGCCACCTCCACTGTGTATGTATGTCGCTCCCTCGTCTGTTCCTCCAGGATC 3557  
Db 5241 CACCACCTGCCACCTCTCCACTGTGTATGTATG---TCCCTGTCTGTTCCTCCACATGAGT 5297  
QY 3558 TCGAAGTGAATCCGGGCTGAGCAGTGGGGCGGCTGGGGAGGGGTG-----ACGATTCT 3611  
Db 5298 GTGACGTGACCGGGGTAGTTGGGAGGGTGCAACTGGAGGGAAGGGCCAGTGCAACTTT 5357  
QY 3612 CCTCAGGCTTTGGCCCTGCAAGCAACCCACATATCTGTCTGTATGTAATAATGCTT 3671  
Db 5358 CCTCAGGCTTTGGCCCTGCAAGTGAACACTCATAATCTGTCTGTATGTAATAATGCTT 5417  
QY 3672 AACGTGTAATAAAAAAAAAAAAA 3694  
Db 5418 AACACTAAAAAAAAAAAAAAAAAAAA 5440

RESULT 8  
AB093282  
LOCUS  
DEFINITION Mus musculus mRNA for mKIAA1066 protein.  
ACCESSION AB093282  
VERSION AB093282.2 GI:28394198  
KEYWORDS FLI\_CDNA.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AB093282 5411 bp mRNA linear ROD 15-FEB-2003  
Mus musculus mRNA for mKIAA1066 protein.  
AB093282  
AB093282.2 GI:28394198  
FLI\_CDNA.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;









```
Db      5058 CCAGCTCCAGACACG-TACTAAGTGCCTAGGTTGCCGGCTGTGGCCTGCTCCCTTGGGA 5116
QY      3204 ----AGCAACAGAGAGGCCCAACAGCAGAGGCCCGTGGGCTGAGGATGGAGCGGCCCCC 3259
Db      5117 GAGCAGCTGACAAGAGGCTTCCAGGCAGAGGCCAGAGAGCTGAGGA-----CCAGCC 5169
QY      3260 AGCCGACTCCAGCCCGCAGAGGGCAGACGCCACCCCTGGACTGCTCTCCCTGCCAGCTG 3319
Db      5170 AGGATGGGCACAGTTCTAAGCAAGTGTGTGGGACAGGTGTGCTCCCTGGCTCTGGCCTTG 5229
QY      3320 GGCTCTCTGGCTATTCTTCTACCTTCCAGGCCCACTGCACTCTGTCTGGGAGGCCCTTA 3379
Db      5230 CTCGTAGGGGC-----A 5242
QY      3380 TGAGGGCAGCCCGCAGCCCGCCGACCCCAACCCAGAGAGACAGATCTTGGGGAGCT 3439
Db      5243 CTGAGTCAGTTGGCCTCTACTGCAGCTTCCCACAGAGAGCAGATCCAGGGCACT 5302
QY      3440 GCCCCACAAGCCCGCTGGCCACCGAGGGCTGCAGC--CGCTGCGCTGCCGGCTTCTCCC 3497
Db      5303 GCCCC-TGAGCCTGTCTTGCCCAACAAGCTGCAGCAGTGTGCAGCTGCAGGCTTCTCCC 5361
QY      3498 CACCACCTGCCACCTCCACTGTGATGTATGTCGCTCCCTCGCTGTCTTCCCCAGGATC 3557
Db      5362 CACCACCTGCCACCTCTCCACTGTGATGTAT---GTCCCTTGTCTGTCTTCCCCACATGAGT 5418
QY      3558 TCGAAGTGACTCCGGGCTGAGCAGTGGGGCGGCTGGGGAGGGGTG-----ACGATTCT 3611
Db      5419 GTGACGTGACCGGGGTAGTTGGGAGGGTGCAACTGGAGGGAAGGGCCAGTGCAACTTT 5478
QY      3612 CCTCAGGCTTTGGCCTGCAAGCAAAACCCACATATCTCTCTGTATGT 3659
Db      5479 CCTCAGGCTTTGGCCTTGCAAGTGAACTCACATATCTCTCTGTATCT 5526
```

```
RESULT 10
HS371H6 20612 bp DNA linear PRI 19-APR-2001
LOCUS Human DNA sequence from clone LA16-371H6 on chromosome 16 Contains
DEFINITION part of the gene for KIAA0516 protein, part of a gene similar to
NDP kinase, ESTs, an STS and 2 CpG islands, complete sequence.
AL031718
ACCESSION AL031718.11 GI:13937341
VERSION HTG; CpG Island; KIAA0516; kinase.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 20612)
AUTHORS Bagguley,C.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
COMMENT On May 3, 2001 this sequence version replaced gi:13277292.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep LA16-371H6 is
```

part of a clone contig from the tip of the short arm of chromosome 16 spanning 2Mb of pl3.3 (Higgs D.R., Flint J., Daniels R., MRC Molecular Haematology Unit, Institute of Molecular Medicine, Oxford (unpublished)), and is from the Los Alamos, flow sorted human Chromosome 16 libraries constructed by Norman Doggett (unpublished). IMPORTANT: This sequence is not the entire insert of clone LA16-371H6 It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. The true left end of clone LA16-371H6 is at 1 in this sequence. The true left end of clone LA16-447E6 is at 20513 in this sequence. The true right end of clone LA16-361A3 is at 3304 in this sequence.

```
FEATURES
    source
        1..20612
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="16"
            /clone="LA16-371H6"
            /clone_lib="LA16"
            join(269..451,529..718,838..957,1032..1112,1213..1280,
            1458..1699,1873..2044,2114..2227,2378..2416,2508..2687,
            2786..2935,3012..3133)
            /gene="c371H6.1"
            join(<269..451,529..718,838..957,1032..1112,1213..1280,
            1458..1699,1873..2044,2114..2227,2378..2416,2508..2687,
            2786..2935,3012..3133)
            /gene="c371H6.1"
            /note="match: cDNAs: Em:AB011088
            match: ESTs: Em:AI815491 Em:AI816436 Em:AA984532
            Em:AA013330
            match: proteins: Tr:O60271 Wp:CE00363 Tr:O15013"
            /codon_start=3
            /evidence=not_experimental
            /product="c371H6.1 (KIAA0516 protein)"
            /protein_id="CAB72318.1"
            /db_xref="GI:6941067"
            /db_xref="GOA:Q9UPT6"
            /db_xref="SWISS-PROT:Q9UPT6"
            /translation="ASDSDYPPGEMFLDSDVNPEDPGADGVLAGITLVGCATRCNVPR
            SNCSSRGDTPVLKQGEVATIANGVNPSQSTEEATEATEVPDGPSPETATLRPG
            PLTEHVFTDPAPTPSSGPQSGENGPEDSSSTRPEPSPGDPTGAGSSAAPTMWIGA
            QNGWLYVHSAVANWKKCLHSIKLKDSVLHLVHKVRLVALADGTLAIHRGEDGGKRL
            LSNYHMLDLGHPHHSIRCMAYVDYRWVCYKKNKHVIOPKTMQIEASAGOGPGEGKRL
            LLASSPASSIRKSTRPYKSFDAHPRRSQVRLAWIGDGVWVSIRLDSLTRLYHAHT
            HQHLQVDVIEPYVSKMLGTGLGFSFVITALLVAGSLRWVGTGNGVVISIPLTETVV
            LHRGQLGLRANKTSPTSGEGARPGGIHVYDDSDRAASEGFIKYCSMAQAQLCFHG
            HRDAVKFVSVPGNVLTNGSVLDSPAEGPGPAAPAEVGGKLRNLVLSGEGEYI
            DFRIGDGEDDETEEGAGDSQVKPVLKSAERSHIIIVQVSYTPE"
            4620..8266
            /note="CpG island"
            /evidence=not_experimental
            complement(join(4958..5075,5190..5302,5379..5480,
            5589..5719))
            /gene="c371H6.2"
            complement(join(4958..5075,5190..5302,5379..5480,
            5589..5719))
            /gene="c371H6.2"
            /note="match: cDNAs: Em:LI6785 Em:S39901 Em:U61287
            Em:AF033377 Em:U89605 Em:M36981 Em:U85511 Em:M91597
            Em:X92956 Em:X92957 Em:U29656 Em:X97900 Em:X97901
            Em:X68193 Em:X58965 Em:AF043543 Em:M55331 Em:AF043542
            Em:M35970 Em:AF045187 Em:M65037 Em:X97899
            match: ESTs: Em:AA999847 Em:W11654 Em:AA387534 Em:W39079
            Em:W48246 Em:W83979 Em:AI525854 Em:AI609754 Em:H83355
            Em:AA276203 Em:W83990 Em:AI356825 Em:AI799023 Em:AA129462
            Em:AI173169 Em:AA781191 Em:AI587608 Em:AA399289
            match: proteins: Sw:P15532 Sw:P15531 Sw:P19804 Sw:Q90380
            Sw:P52174 Sw:P52175 Tr:O57535 Sw:Q05982 Sw:Q13232
            Sw:P27950 Tr:P70071 Sw:P26550"
            /codon_start=3
            /evidence=not_experimental
            /product="c371H6.2 (similar to NDP kinase)"
```

```
/protein_id="CAB72319.1"
/db_xref="GI:6941068"
/db_xref="GOA:Q9NUF9"
/db_xref="SPTREMBL:Q9NUF9"
/translation="CTGAHRTFLAVKPDGVQRRLLVGEIVRRFERKGFKLVALKLVA
SEELRHYAELRERPFYGLVKYMASGPVAVMQGLDVVRTSRALIGATNPADAPP
GTIRGDFCIKLVGNLIHGSDSVESARREIALWFRADLLCWEDSAGHWLYE"
5536..5589
/note="27 copies 2 mer cc 70% conserved"
6068..6125
/note="29 copies 2 mer cc 69% conserved"
13032..13462
/note="L1M3e repeat: matches 426..872 of consensus"
complement(15564..16020)
/note="match: STS: Em:G37311"
16033..18297
/note="CpG island"
/evidence=not_experimental
16198..16313
/note="58 copies 2 mer cc 59% conserved"

ORIGIN

Query Match
Best Local Similarity 47.4%; Score 1751.4; DB 9; Length 20612;
Matches 1946; Conservative 0; Mismatches 1; Indels 174; Gaps 2;

QY 1734 AGCCAATAAGACATCCCCACCTTGGGGAGGGGCGCCGTCCTGATCCCTACTATCCCTACTGCTCCATGGC 1793
Db |||||
2506 AGCCAATAAGACATCCCCACCTTGGGGAGGGGCGCCGTCCTGATCCCTACTATCCCTACTGCTCCATGGC 2565
QY 1794 GTATGGCGATGACAGCAGTGACAGGGCGGCCAGCAGCTTCAATCCCTACTGCTCCATGGC 1853
Db |||||
2566 GTATGGCGATGACAGCAGTGACAGGGCGGCCAGCAGCTTCAATCCCTACTGCTCCATGGC 2625
QY 1854 CCAGGCCCAGCTATGCTTCCATGGGCACCGGATGCCGTGAAGTTCTTTGTCTGGTG-- 1911
Db |||||
2626 CCAGGCCCAGCTATGCTTCCATGGGCACCGGATGCCGTGAAGTTCTTTGTCTGGTGCC 2685
QY 1912 ----- 1911
Db AGGTGAGGCTGGGCCCCCTCTGCCATCCACATCCCCCTGATGCCAGTGGCGCGCCCTCC 2745
QY 1912 -----CCAGGGAACGTGTGGCCACCCCTG 1935
Db |||||
2746 CCCAGGAGCGCGTGTCTGTAATCGCTTCTGCCATCCAGGGAACGTGTGGCCACCCCTG 2805
QY 1936 AATGGCAGTGTGTGGACAGCCAGCCAGGGCCCTGGGCCAGCTGCCCTGCCCTCGGAG 1995
Db |||||
2806 AATGGCAGTGTGTGGACAGCCAGCCAGGGCCCTGGGCCAGCTGCCCTGCCCTCGGAG 2865
QY 1996 GTCGAGGCCAGAAAGCTGCGGAACGTGTGTGTGCTGAGCGGGGGAGGGCTACATCGAC 2055
Db |||||
2866 GTCGAGGCCAGAAAGCTGCGGAACGTGTGTGTGCTGAGCGGGGGAGGGCTACATCGAC 2925
QY 2056 TTCCGCATT----- 2064
Db TTCCGCATTGGTGAGCGGGGCCCCAGGACAGGGCTGAGGTTGGGCGCGGGGGAGCCTGG 2985
QY 2065 -----GGAGACGAGAGGACGACGAGACGAGACGAGAGGAGGAGGCGC 2099
Db |||||
2986 CCCTACTGTGCTGTTTGGCCCGCAGGAGACGAGAGGAGGACGAGACGAGAGGAGGCGC 3045
QY 2100 AGGGGACATGAGCCAGGTGAAGCCCGTGTGTCCAAGGACAGCGCAGTCACATCATCGT 2159
Db |||||
3046 AGGGGACATGAGCCAGGTGAAGCCCGTGTGTCCAAGGACAGCGCAGTCACATCATCGT 3105
QY 2160 GTGGCAGGTGTCTTACACCCCGAGTGAAGCTGTGCCCTGCTGGCCGACCTGTACAT 2219
Db |||||
3106 GTGGCAGGTGTCTTACACCCCGAGTGAAGCTGTGCCCTGCTGGCCGACCTGTACAT 3165
QY 2220 AGGACCCCGACCACTGACCCCGCCCGCGGGGTAGCCAGCCAGGCGCCGCCGC 2279
Db |||||
3166 AGGACCCCGACCACTGACCCCGCCCGCGGGGTAGCCAGCCAGGCGCCGCCGC 3225
```

```
QY 2280 CCCTCTTCTAAACCTCTCAACCTGACGCTTTCACTGAGTCTTGCCCTCCAGCGGGCAGG 2339
Db |||||
3226 CCCTCTTCTAAACCTCTCAACCTGACGCTTTCACTGAGTCTTGCCCTCCAGCGGGCAGG 3285
QY 2340 GAGTGGGGGATGCGGATCAGCTGGGAGGAGGAGGGGGTGTCTTCCACCCGAGGGA 2399
Db |||||
3286 GAGTGGGGGATGCGGATCAGCTGGGAGGAGGAGGGGTGTCTTCCACCCGAGGGA 3345
QY 2400 AGATGCTCTCGGACAGTTTCCCGGCGAGCTCTTGCCAGCTTCCAGCCAGAGTCTCA 2459
Db |||||
3346 AGATGCTCTCGGACAGTTTCCCGGCGAGCTCTTGCCAGCTTCCAGCCAGAGTCTCA 3405
QY 2460 AGTCCAGGGCACCTTGGGGCCAGCGCAGGAGGAGTCCGAGGTGTCTTGGCTCTTACCTG 2519
Db |||||
3406 AGTCCAGGGCACCTTGGGGCCAGCGCAGGAGGAGTCCGAGGTGTCTTGGCTCTTACCTG 3465
QY 2520 GGCCTCTACTCTCCAGCACCTTGGGAGGAGGAGGGGCTCCCGCGCGAGGCTGCCT 2579
Db |||||
3466 GGCCTCTACTCTCCAGCACCTTGGGAGGAGGAGGGGCTCCCGCGCGAGGCTGCCT 3525
QY 2580 GGCCTGGGCCACCTCTGCTGCTGCTCATGGGGCCACCTTGGCTCTTGGGGCTTCACTC 2639
Db |||||
3526 GGCCTGGGCCACCTCTGCTGCTGCTCATGGGGCCACCTTGGCTCTTGGGGCTTCACTC 3585
QY 2640 TGCCTAGGGAGTGGGCCAGGCACTAGCCTTTTGGCCAGGGAGTGGGCTCAGGCTGCC 2699
Db |||||
3586 TGCCTAGGGAGTGGGCCAGGCACTAGCCTTTTGGCCAGGGAGTGGGCTCAGGCTGCC 3645
QY 2700 CAGGTGCTGCACCCAGCCAGCGGCTTCTCTGGGGCTTCCCGCTGCTCAAGCTCTATCCT 2759
Db |||||
3646 CAGGTGCTGCACCCAGCGGCTTCTCTGGGGCTTCCCGCTGCTCAAGCTCTATCCT 3705
QY 2760 GTCTGTCCCAACCCAGCTGTCCCTGCCCCAGGAGCTGGCATAAAGCACAGGCCCCGG 2819
Db |||||
3706 GTCTGTCCCAACCCAGCTGTCCCTGCCCCAGGAACTGGCATAAAGCACAGGCCCCGG 3765
QY 2820 CTCCCTGGGCGAGCTGCTTGAGAACAGAGACTGTACCCCATCTGCCCATGCGAGGCAGG 2879
Db |||||
3766 CTCCCTGGGCGAGCTGCTTGAGAACAGAGACTGTACCCCATCTGCCCATGCGAGGCAGG 3825
QY 2880 CTCTTGCCAGCCCGCTTCTGACCCCGTGTCCCGCCAGGCTCTGCTGGGAGAAAGACTCAC 2939
Db |||||
3826 CTCTTGCCAGCCCGCTTCTGACCCCGTGTCCCGCCAGGCTCTGCTGGGAGAAAGACTCAC 3885
QY 2940 CTTGAGGAGTGGGCGCTGAGTCTGCTCCCTCCAGAAAGCCCGGAGGTGGGATTTCTC 2999
Db |||||
3886 CTTGAGGAGTGGGCGCTGAGTCTGCTCCCTCCAGAAAGCCCGGAGGTGGGATTTCTC 3945
QY 3000 AGGCTGCCAGGGCAGGCCAGGCTCAGGAAAGAGGGAGGCGCCCTGCGCTCTCCGGGAT 3059
Db |||||
3946 AGGCTGCCAGGGCAGGCCAGGCTCAGGAAAGAGGGAGGCGCCCTGCGCTCTCCGGGAT 4005
QY 3060 CAGTCTTAGGACACAGGCTCAGCCTCAGGTTGATGGGGATGATGTCTCCCGGGCCTG 3119
Db |||||
4006 CAGTCTTAGGACACAGGCTCAGCCTCAGGTTGATGGGGATGATGTCTCCCGGGCCTG 4065
QY 3120 CCTCTGACGGGGCTCCAGGAGCCAGCTCCAGACACGCTACTAAGTGCCTAGGGTT 3179
Db |||||
4066 CCTCTGACGGGGCTCCAGGAGCCAGCTCCAGACACGCTACTAAGTGCCTAGGGTT 4125
QY 3180 GCGCGCTGTGGCTGCTCCAGGGAGCAACAGAGAGGCTCCAGGAGGCGCCCTGGGG 3239
Db |||||
4126 GCGCGCTGTGGCTGCTCCAGGGAGCAACAGAGAGGCTCCAGGAGGCGCCCTGGGG 4185
QY 3240 CTGAGGATGGAGCGCGCCAGCCGACTCCAAGCCCGAGAGGGGAGCGCCACCTTGA 3299
Db |||||
4186 CTGAGGATGGAGCGCGCCAGCCGACTCCAAGCCCGAGAGGGGAGCGCCACCTTGA 4245
QY 3300 CTGCTCTCCCTGCGCAGCTGGGCTTCTTGGCTTATCTTACCTTCCAGGCGCCCTGAC 3359
Db |||||
4246 CTGCTCTCCCTGCGCAGCTGGGCTTCTTGGCTTATCTTACCTTCCAGGCGCCCTGAC 4305
```

QY 3360 TCCTGTCTGGAGGCCCTTATGAGGGCAGCCAGCCCCCGCACCCACCCCGCCCAACCAGAGA 3419  
|||||  
Db 4306 TCCTGTCTGGAGGCCCTTATGAGGGCAGCCAGCCCCCGCACCCACCCCGCCCAACCAGAGA 4365  
|||||  
QY 3420 AGCACAGATCTTGGGAGCTGCCACAAAGCCCGCTGGCCACGAGGGCTGCAGCCGCT 3479  
|||||  
Db 4366 AGCACAGATCTTGGGAGCTGCCACAAAGCCCGCTGGCCACGAGGGCTGCAGCCGCT 4425  
|||||  
QY 3480 GCGCTGCCGGCTTCTCCCCACCAACCCCTGCCACTCCACTGTGATGTATGTCCGCTCCCTC 3539  
|||||  
Db 4426 GCGCTGCCGGCTTCTCCCCACCAACCCCTGCCACTCCACTGTGATGTATGTCCGCTCCCTC 4485  
|||||  
QY 3540 GTCTGTTCCCCCAGGATCTCGAAGTGACTCCGGCTGAGCAGTGGGCGGCTGGGGGAGG 3599  
|||||  
Db 4486 GTCTGTTCCCCCAGGATCTCGAAGTGACTCCGGCTGAGCAGTGGGCGGCTGGGGGAGG 4545  
|||||  
QY 3600 GGTGACGATTCTCCTCAGGCTTTGGCCCTGCAAGCAAAACCCACATATCTGCTGTATGT 3659  
|||||  
Db 4546 GGTGACGATTCTCCTCAGGCTTTGGCCCTGCAAGCAAAACCCACATATCTGCTGTATGT 4605  
|||||  
QY 3660 AATAAATGCTTAACGTCGTA 3680  
|||||  
Db 4606 AATAAATGCTTAACGTCGTA 4626  
|||||

RESULT 11  
AE006639  
LOCUS AE006639 270150 bp DNA linear PRI 15-AUG-2002  
DEFINITION Homo sapiens 16p13.3 sequence section 7 of 8.  
ACCESSION AE006639 AE005175  
VERSION AE006639.1 GI:14336758  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 270150)  
AUTHORS Daniels,R.J., Peden,J.F., Lloyd,C., Horsley,S.W., Clark,K.,  
Tufarelli,C., Kearney,L., Buckle,V.J., Doggett,N.A., Flint,J. and  
Higgs,D.R.  
TITLE Sequence, structure and pathology of the fully annotated terminal 2  
Mb of the short arm of human chromosome 16  
JOURNAL Hum. Mol. Genet. 10 (4), 339-352 (2001)  
MEDLINE 21096910  
PUBMED 11157797  
REFERENCE 2 (bases 1 to 270150)  
AUTHORS Daniels,R.J., Peden,J.F., Lloyd,C., Horsley,S.W., Clark,K.,  
Tufarelli,C., Kearney,L., Buckle,V.J., Doggett,N.A., Flint,J. and  
Higgs,D.R.  
TITLE Direct Submission  
JOURNAL Submitted (07-DEC-2000) MRC Molecular Haematology Unit, Weatherall  
Institute of Molecular Medicine, John Radcliffe Hospital, Oxford,  
Oxon OX3 9DS, UK  
FEATURES  
source Location/Qualifiers  
1..270150  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/map="16p13.3"  
1..5613  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="cosmid 425c2"  
1..37670  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
59799..93949  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"

source /db\_xref="taxon:9606"  
/clone="cosmid 431h6"  
89167..121683  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="cosmid 329f2"  
113820..160576  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="cosmid 361a3"  
157079..193613  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="cosmid 371h6"  
<1..25277  
/note="HS395F10"  
complement(54..172)  
/note="FRAM; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="SINE/Alu"  
190..477  
/note="AluSg; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="SINE/Alu"  
1371..4511  
/note="CpG island; KIAA0590"  
/evidence=not experimental  
complement(1479..4124)  
/gene="gs114"  
complement(join(1479..1729,1981..2170,3687..4124))  
/gene="gs114"  
/note="GENSCAN prediction; no supporting evidence known at  
this time"  
/codon\_start=1  
/product="unknown"  
/protein\_id="AAK61287.1"  
/db\_xref="GI:14336759"  
/translation="MGTKPPRPEASCCTGRCGSDAPPALSRAARQDPPAGPRHDS  
AAQWSPSSTPGSLASRGLPGCPQNPGRAPAPERRIQGRNAGLVVRLGSDWSS  
GASPGTSPRAGARLGRITQILPPLTLAKPARRPQQFRLRRPSGITWPTLGP  
TCVPDRLLFPAAARGASVLGAASLTFGYFRVPGAPLSQRAAAQRPGRAGRVPLEWA  
GWGGRRHHTPEAKRDKEMQNKNMCLVTRWVGPEGRVNASRQRMASQYRPLDAART  
LRRPKALDCGLEAL"  
2542..2836  
/note="AluJb; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="SINE/Alu"  
2849..3156  
/note="AluSx; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="SINE/Alu"  
3606..3663  
/note="(CCCCCG)n; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="Simple\_repeat"  
3766..3793  
/note="GC rich; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="Low\_complexity"  
3874..3908  
/note="GC rich; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="Low\_complexity"  
5288..69224  
/gene="KIAA1426"  
/note="synonym: CRAMP 1 like"  
join(5288..5397,5955..6301,11954..12007,17352..17482,  
23528..23681,29140..29223,32455..32503,34063..34273,  
35284..35356,35558..35697,43780..43897,44349..44472,  
44552..44707,46535..46607,47142..48308,49423..49620,

```
CDS
51202. .51379,53806. .53987,54077. .54151,56373. .56454,
57389. .57493,57738. .57916,58651. .58716,59278. .59509,
60317. .60481,61980. .62125,65197. .69222)
/gene="KIAA1426"
/note="Similar to Unigene clusters HS.12849 and
HS.117900"
join(5288. .5397,5955. .6301,11954. .12007,17352. .17482,
23528. .23681,29140. .29223,32455. .32503,34063. .34273,
35284. .35356,35558. .35697,43780. .43897,44349. .44472,
44552. .44707,46535. .46607,47142. .48308,49423. .49620,
51202. .51379,53806. .53987,54077. .54151,56373. .56454,
57389. .57493,57738. .57916,58651. .58716,59278. .59509,
60317. .60481,61980. .62125,65197. .65361)
/gene="KIAA1426"
/note="Peptide sequence identity with GP:Y13674_1;
Structure of CDS predicted by GENSCAN/genewise; last
coding exon was not predicted by GENSCAN but supported by
cDNA emb1:AB037847"
/product="unknown"
/protein_id="AAK61288.1"
/db_xref="GI:14336760"
/translation="MEFPVRAGEPALLPSRGGSLCPPPGALGAGAWGRMTVKLGD
GGGEDGLKLGKRAADEESLEGEAGGADAAEESGTRDEKTPRAGADGPPAPGA
PQAPSPQGS PQDQHFLRSSVR PQSKRPKDPSPSAVSGNAGSGPRGKDRHLPKAQ
LWLSIRVPFGSSSSRNLSGSGEKEEGKVRQWESWSTEDKNTFFEGLYEHGKD
FEAIQNNIALKYKKKGKSPASVMKNKEQVHFYRTWHKITKYIDFDHVFSGRLKSSQ
ELYGLICYGELRRKIGGMDKNATKLNELIQVHTGWGSKYFTFYFIS MIDGKKPE
FQTLCCMLEDGAQSLDSERSFCQNTDVLPSGGVVGTCSAIRGRTYEASVEVAHLATS
LSGDWAGGKRKERCETRRKLEOHTSKERFLSAPMERSEAEILLNSVFIGSGNVWV
PASAFLONSFPRA TVRYKGRNLR IKAPCMRALKKLCDDPDGLSDEEDQKPVRLPIKVP
IELQPRNNHAWARVQSLAONPRLNFOEQVHPVYALSSHEDAAVWRRLRESREHAAVL
YLGRDPTCVQAVEGMSRMI VELHRKVSSLI EFLKQKWLHEHPDL SASQCPSLTGT
QRKLEERQLQDSCSAPMOEKVTLHLFPGENCTLTPLPGVARVVHSAFCTVHWQEGG
RCKQSAKADHVLPPAQLIGIQSGGTARGQVKCPRSGAEGKGVGRPPPAADALQSSGE
SSPESAPGEAALSLSPPADPRPPRHQDTGPCLEKTPAEGRDSPTREP GALPCACG
QLPDLDEL SLLDPLRYLKSQDLIVPEQCRADTRPGEQPLGGGAASPEVLA PVS
KEAADLAPTGPSRPGPGLLDVCTKDADAPAELEQKSGPAGPPPSQGPAPRPPK
EVPASRLAQLREEGWNLTSESLTAEVYLMGKPSKLQLEYDWLGPGRQDPRPSL
PTALHKORLLSCLLKLISTEVPNPKLPVQGLPRRVGPSAGGRNVQICLSIKAGGKKV
LPGVVETEPNQA PLRPSLRVKVARGSHSRDTRALEANTIISTASVRPAQEESQMTPPG
KVTVSSRSRPRCPRNQASLRSSKTFPPSSAPCSGLRNPRLPVGPSSTGSNDS DG
GLFAVPTTLPPNSRHGKLFSPSKEAELTFRQHLNLSISMQSDFFLPKPKRLNRHLRKP
LVVQRTLLPRPSENQSHNVCSFILSNSVTVGRGSFRPIQSSLTKAALSRIPIVKVLP
PQATSHLASAIDLAA TSAGILSGNPLPALDTEGLSGISPLSSDEVTGAI SGDS TGTH
QDQDTLPTVGGSDPFVSI PSRPEQEPVADSFQSSVLSLSELPKAPLQNGLSIPLSSS
ESSSTRLSPPDVSA LLDISLPGPPEDALSQGEPA THISDSIIEIAISSGQYGEVPLS
PAKLNGSDSSKSLPSPSSSQPHWIASPTHDPQWYPSDSTSSLSLFA SFISPEKSR
XMLPTPIGNTSGTSLGPSLLDGNRSDSFVRSRLADVAEVVDSQLVCMNNENSIDYIS
RFNDLAQELSIAEPGRREALFDGGGGPAVSDLSQ"
5511. .6410
/gene="KIAA1426"
/note="CpG island; CRAMP1L"
/evidence=not_experimental
5625. .5700
/note="GC rich; RepeatMasker predicted
23/10/2000"
/rpt_family="Low_complexity"
5722. .5828
/note="G-rich; RepeatMasker predicted
23/10/2000"
/rpt_family="Simple_repeat"
6106. .6171
/note="GC rich; RepeatMasker predicted
23/10/2000"
/rpt_family="Low_complexity"
7716. .7810
/note="MIR; RepeatMasker predicted
23/10/2000"
misc_feature
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
```

```
repeat_region
/rpt_family="SINE/MIR"
complement(8132. .8438)
/note="AluY; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
9506. .9625
/note="AluSp/q; RepeatMasker predicted
23/10/2000"
repeat_region
Query Match 47.4%; Score 1751.4; DB 9; Length 270150;
Best Local Similarity 91.7%; Pred. No. 1.7e-247;
Matches 1946; Conservative 0; Mismatches 1; Indels 174; Gaps 2;
QY 1734 AGCCAATAAGACATCCCCACCTCTGGGAGGCGCCCGTCCGGGGGCATCATCCACGT 1793
Db |||||||
159584 AGCCAATAAGACATCCCCACCTCTGGGAGGCGCCCGTCCGGGGGCATCATCCACGT 159643
QY 1794 GTATGGCGATGACAGCAGTGTGACAGGGCGGCAGCAGCTTCAATCCCTACTGTCTCCATGGC 1853
Db |||||||
159644 GTATGGCGATGACAGCAGTGTGACAGGGCGGCAGCAGCTTCAATCCCTACTGTCTCCATGGC 159703
QY 1854 CCAGGCCACGCTATGCTTCCATGGGCACCGCGATGCCGATGCCGATGCGGTGCTCGGTG-- 1911
Db |||||||
159704 CCAGGCCACGCTATGCTTCCATGGGCACCGCGATGCCGATGCGGTGCTCGGTGCTCGGTGCC 159763
QY 1912 -----
Db 159764 AGGTGAGGTGGGCCCTCCTCGCATCCACATCCCCTGATGCCAGTGGCGCGCCCTCC 159823
QY 1912 -----CCAGGGAACGTGCTGGCCACCCCTG
159824 CCAGGAGCGCGTGTCTCTGAATCGTTCTGCCATCCCAGGGAACGTGCTGGCCACCCCTG 159883
QY 1936 AATGGCAGTGTGTGGACACCCAGCCAGCGGCGCCCTGGGCCAGTGCCTGCTCGGTGGAG 1995
Db |||||||
159884 AATGGCAGTGTGTGGACACCCAGCCAGCGGCGCCCTGGGCCAGTGCCTGCTCGGTGGAG 159943
QY 1996 GTCAGGGCCAGAAGCTGCGGAACGTGCTGCTGTAGCGCGGGGAGGGGTACATCGAC 2055
Db |||||||
159944 GTCAGGGCCAGAAGCTGCGGAACGTGCTGCTGTAGCGCGGGGAGGGGTACATCGAC 160003
QY 2056 TTCCCGCAATT-----
Db 160004 TTCCCGCAFTGGTGAGCGGGGCCAGGGACAGGGCTGAGTTGGGCGCGGGGGAGCCTGG 160063
QY 2065 -----GGAGACGGAGAGGACGACGAGACGGAGGAGGGCGC
160064 CCTCACTCTGCTGTTTGGCCCGCAGGAGACGGAGACGAGACGGAGGAGGGCGC 160123
QY 2100 AGGGACATGAGCCAGGTGAAGCCCGTGTGTCCAAGCAGAGCGCAGTCACTCATCGT 2159
Db |||||||
160124 AGGGACATGAGCCAGGTGAAGCCCGTGTGTCCAAGCAGAGCGCAGTCACTCATCGT 160183
QY 2160 GTGCAGGTGTCTTACACCCCGAGTGAAGCTGTGCTGCCCTGCCCTGCCACCTGTATCAT 2219
Db |||||||
160184 GTGCAGGTGTCTTACACCCCGAGTGAAGCTGTGCTGCCCTGCCCTGCCACCTGTATCAT 160243
QY 2220 AGGACCCCGACCACTGACCCCGCCCGCGCGCGGGGTAGCCAGCAGGCGCGCGCGC 2279
Db |||||||
160244 AGGACCCCGACCACTGACCCCGCCCGCGCGCGGGGTAGCCAGCAGGCGCGCGCGC 160303
QY 2280 CCCTCTTCTAACCTCTCAACCTGCAGCTTTCACCTGAGTCTGGCCCTCCAGCGGGCAGG 2339
Db |||||||
160304 CCCTCTTCTAACCTCTCAACCTGCAGCTTTCACCTGAGTCTGGCCCTCCAGCGGGCAGG 160363
QY 2340 GAGTCGGGGATGCGGATCAGCTGGAGGAGGAGGGGTGCTTCCACCCGAGGGGA 2399
Db |||||||
160364 GAGTCGGGGATGCGGATCAGCTGGAGGAGGAGGGGTGCTTCCACCCGAGGGGA 160423
QY 2400 AGATGCTCTCGGACAGTTTCCCGGGCAGCTCTCGGCCAGCTTCCAGCCCGAGTCTCTCA 2459
Db |||||||
160424 AGATGCTCTCGGACAGTTTCCCGGGCAGCTCTCGGCCAGCTTCCAGCCCGAGTCTCTCA 160483
```

|    |        |   |        |  |
|----|--------|---|--------|--|
| QY | 2460   | AGTCCAGGACACCTTGGGCCAGCGCAGGAGAGTCCGAGGTGGTCTCTGGCTACCCCTG  | 2519   |  |
| Db | 160484 | AGTCCAGGACACCTTGGGCCAGCGCAGGAGAGTCCGAGGTGGTCTCTGGCTACCCCTG  | 160543 |  |
| QY | 2520   | GGCCTCCTACTCCCCAGCACCCCTGGAGGAGGAGGGCTCCCGCCGCGGAGGCTGCCT   | 2579   |  |
| Db | 160544 | GGCCTCCTACTCCCCAGCACCCCTGGAGGAGGAGGGCTCCCGCCGCGGAGGCTGCCT   | 160603 |  |
| QY | 2580   | GGCCTGGGCCACCTCTGCATGCTCTATGGGGCCACCCCTGCTCTGGGCCCTCACTC    | 2639   |  |
| Db | 160604 | GGCCTGGGCCACCTCTGCATGCTCTATGGGGCCACCCCTGCTCTGGGCCCTCACTC    | 160663 |  |
| QY | 2640   | TGCCTAGGGAGCTGGGCCAGGCACTAGCCCTTGTCCCGAGGAGGTGGGCTCAGGCTGCC | 2699   |  |
| Db | 160664 | TGCCTAGGGAGCTGGGCCAGGCACTAGCCCTTGTCCCGAGGAGGTGGGCTCAGGCTGCC | 160723 |  |
| QY | 2700   | CAGGTGCCTGCACCCCGAGCCGCTTCTCTGGGGCCTCCCGCTCAAGCCTCTATCCT    | 2759   |  |
| Db | 160724 | CAGGTGCCTGCACCCCGAGCCGCTTCTCTGGGGCCTCCCGCTCAAGCCTCTATCCT    | 160783 |  |
| QY | 2760   | GTCTGTCCCCACCCAGCTGTCCCTGTCCCGAGGAGCTGGCAFAAAGACAGGCGCCCG   | 2819   |  |
| Db | 160784 | GTCTGTCCCCACCCAGCTGTCCCTGTCCCGAGGAGCTGGCAFAAAGACAGGCGCCCG   | 160843 |  |
| QY | 2820   | CTCCCTGGGGAGCTGCTTGAGAACAGAGACTGTACCCCATCTGCTGCCATGAGGCAGG  | 2879   |  |
| Db | 160844 | CTCCCTGGGGAGCTGCTTGAGAACAGAGACTGTACCCCATCTGCTGCCATGAGGCAGG  | 160903 |  |
| QY | 2880   | CTCTTGCCAGCCCGTCTGTACCCGTGTCCCGCAGGCTCTGCTGGGCAGAACTCAC     | 2939   |  |
| Db | 160904 | CTCTTGCCAGCCCGTCTGTACCCGTGTCCCGCAGGCTCTGCTGGGCAGAACTCAC     | 160963 |  |
| QY | 2940   | CTTGAGGAGTGGGCTGGAGTCTGCTCCCTCCAGAAAGCCCGAGGGTGGGATTCTC     | 2999   |  |
| Db | 160964 | CTTGAGGAGTGGGCTGGAGTCTGCTCCCTCCAGAAAGCCCGAGGGTGGGATTCTC     | 161023 |  |
| QY | 3000   | AGGCTGCCAGGGCAGGCCAGGCTCAGGAAGAGGGAGGCCCTCTCCGGGAT          | 3059   |  |
| Db | 161024 | AGGCTGCCAGGGCAGGCCAGGCTCAGGAAGAGGGAGGCCCTCTCCGGGAT          | 161083 |  |
| QY | 3060   | CAGTCTTAGGACACAGGCTCAGCCTCAGGTTGATGGGGATGATGCTCCCGGGCCTG    | 3119   |  |
| Db | 161084 | CAGTCTTAGGACACAGGCTCAGCCTCAGGTTGATGGGGATGATGCTCCCGGGCCTG    | 161143 |  |
| QY | 3120   | CCTCTGCACGGGGCTCCAGGAGCCAGCTCCCGACACAGCTACTAAGTGCCTAGGGTT   | 3179   |  |
| Db | 161144 | CCTCTGCACGGGGCTCCAGGAGCCAGCTCCCGACACAGCTACTAAGTGCCTAGGGTT   | 161203 |  |
| QY | 3180   | GCCCGTGTGGCCTGTCCAGGGAGCAACAGAGAGGCCACCAAGCAGAGGCCCGTGGG    | 3239   |  |
| Db | 161204 | GCCCGTGTGGCCTGTCCAGGGAGCAACAGAGAGGCCACCAAGCAGAGGCCCGTGGG    | 161263 |  |
| QY | 3240   | CTGAGATGGAGCCGCCAGCCGCTCAAGCCGAGAGGGCAGACGCCACCTGGA         | 3299   |  |
| Db | 161264 | CTGAGATGGAGCCGCCAGCCGCTCAAGCCGAGAGGGCAGACGCCACCTGGA         | 161323 |  |
| QY | 3300   | CTGCTCTCCCTGCCAGCTGGGCTCTCTGGCCTATTCTACCTTCCAGGCCCACTGCAC   | 3359   |  |
| Db | 161324 | CTGCTCTCCCTGCCAGCTGGGCTCTCTGGCCTATTCTACCTTCCAGGCCCACTGCAC   | 161383 |  |
| QY | 3360   | TCCTGTCTGGGAGCCCTTATGAGGGCAGCCCGCAGCCCGCAGCCCGCAGAGAGAG     | 3419   |  |
| Db | 161384 | TCCTGTCTGGGAGCCCTTATGAGGGCAGCCCGCAGCCCGCAGCCCGCAGAGAGAG     | 161443 |  |
| QY | 3420   | AGCACAGATCTTGGGGAGTGCCCCCAAGCCCGCTGCGCCAGGAGGCTGCAGCGCT     | 3479   |  |
| Db | 161444 | AGCACAGATCTTGGGGAGTGCCCCCAAGCCCGCTGCGCCAGGAGGCTGCAGCGCT     | 161503 |  |
| QY | 3480   | GGCTGCGGCTTCTCCCGACACCCCTGCACCTCCACTGTGATGTATGCTCCGCTCCCTC  | 3539   |  |
| Db | 161504 | GGCTGCGGCTTCTCCCGACACCCCTGCACCTCCACTGTGATGTATGCTCCGCTCCCTC  | 161563 |  |
| QY | 3540   | GTCTGTTCCCCAGGATCTCGAAGTACTCCGGGCTGAGCAGTGGGGCGGCTGGGGAGG   | 3599   |  |

|                       |  |   |                 |               |
|-----------------------|--|---|-----------------|---------------|
| Db                    | 161564   | GTCTGTTCCCCAGGATCTCGAAGTGA                                      | 161623          |               |
| QY                    | 3600   | GGTGACGATTCCTCAGGCTTGGCCCTGCAAGCAAAACCATATCTGCTCTGTATGT         | 3659            |               |
| Db                    | 161624   | GGTGACGATTCCTCAGGCTTGGCCCTGCAAGCAAAACCATATCTGCTCTGTATGT         | 161683          |               |
| QY                    | 3660   | AATAAATGTCTTAACGTCGTA   | 3680            |               |
| Db                    | 161684   | AATAAATGTCTTAACGTCGTA   | 161704          |               |
| RESULT 12             |  |   |                 |               |
| AC012180              |  |   |                 |               |
| LOCUS                 | AC012180   | Homo sapiens chromosome 16 clone RP11-31110, linear             | PRI 20-SEP-2002 |               |
| DEFINITION            | AC012180   | Homo sapiens chromosome 16 clone RP11-31110, complete sequence. |                 |               |
| ACCESSION             | AC012180   |   |                 |               |
| VERSION               | AC012180.6   | GI:23237941   |                 |               |
| KEYWORDS              | HTG.   |   |                 |               |
| SOURCE                | Homo sapiens (human)   |   |                 |               |
| ORGANISM              | Homo sapiens   |   |                 |               |
|                       | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  |   |                 |               |
|                       | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.         |   |                 |               |
| REFERENCE             | 1 (bases 1 to 69437)   |   |                 |               |
| AUTHORS               | DOE Joint Genome Institute, Stanford Human Genome Center and Los   |   |                 |               |
|                       | Alamos National Laboratory.  |   |                 |               |
| TITLE                 | Direct Submission  |   |                 |               |
| JOURNAL               | Unpublished  |   |                 |               |
| REFERENCE             | 2 (bases 1 to 69437)   |   |                 |               |
| AUTHORS               | DOE Joint Genome Institute.  |   |                 |               |
| TITLE                 | Direct Submission  |   |                 |               |
| JOURNAL               | Submitted (21-OCT-1999) production Sequencing Facility, DOE Joint  |   |                 |               |
|                       | Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA |   |                 |               |
| REFERENCE             | 3 (bases 1 to 69437)   |   |                 |               |
| AUTHORS               | DOE Joint Genome Institute, Stanford Human Genome Center and Los   |   |                 |               |
|                       | Alamos National Laboratory.  |   |                 |               |
| TITLE                 | Direct Submission  |   |                 |               |
| JOURNAL               | Submitted (20-SEP-2002) DOE Joint Genome Institute, 2800 Mitchell  |   |                 |               |
|                       | Drive, Walnut Creek, CA 94598, USA                                 |   |                 |               |
| COMMENT               | On Sep 20, 2002 this sequence version replaced gi:13786344.        |   |                 |               |
|                       | Draft Sequence Produced by DOE Joint Genome Institute              |   |                 |               |
|                       | www.jgi.doe.gov  |   |                 |               |
|                       | Finishing Completed at Stanford Human Genome Center and Los Alamos |   |                 |               |
|                       | National Laboratory  |   |                 |               |
|                       | www.shgc.stanford.edu  |   |                 |               |
|                       | Quality: Phrap Quality >=40 99.7% of Sequence;                     |   |                 |               |
|                       | Estimated Total Number of Errors is 0.3.                           |   |                 |               |
|                       | NOTE: This is not the entire sequence of the clone (entire         |   |                 |               |
|                       | sequence is 170.3kb). It is clipped at the overlap with AL031724.  |   |                 |               |
|                       | The number of bases overlapped is 46229.                           |   |                 |               |
| FEATURES              | Location/Qualifiers  |   |                 |               |
| source                | 1..69437   |   |                 |               |
|                       | /organism="Homo sapiens"   |   |                 |               |
|                       | /mol_type="genomic DNA"  |   |                 |               |
|                       | /db_xref="taxon:9606"  |   |                 |               |
|                       | /chromosome="16"   |   |                 |               |
|                       | /clone="RP11-31110"  |   |                 |               |
| ORIGIN                |  |   |                 |               |
| Query Match           | 47.3%;   | Score 1746.6;   | DB 9;           | Length 69437; |
| Best Local Similarity | 91.6%;   | Pred. No. 1.3e-246;   |                 |               |
| Matches 1943;         | Conservative   | 0;  | Mismatches      | 4;            |
|                       |  |   | Indels          | 174;          |
|                       |  |   | Gaps            | 2;            |
| QY                    | 1734   | AGCCATAAGACATCCCCACCTCTTGGGAGGGCGCCCGTCCCGGGGCATCATCCACGT       | 1793            |               |
| Db                    | 5208   | AGCCATAAGACATCCCCACCTCTTGGGAGGGCGCCCGTCCCGGGGCATCATCCACGT       | 5267            |               |
| QY                    | 1794   | GTATGGGATGACAGCAGTGACAGGGCGCCAGAGTTCATCCCTACTGCTCCATGGC         | 1853            |               |
| Db                    | 5268   | GTATGGGATGACAGCAGTGACAGGGCGCCAGAGTTCATCCCTACTGCTCCATGGC         | 5327            |               |
| QY                    | 1854   | CCAGGCCAGCTATGCTTCCATGGGACCGCGATGCCGTGAAGTCTTTGTCTCGGTG--       | 1911            |               |

Db 5328 CCAGGCCAGCTATGCTTCCATGGGCAACCGGATGCCGTGAAGTTCTTTGTCTCGGTGCC 5387  
Qy 1912 ----- 1911  
Db 5388 AGGTAGGCTGGCCCTCCTGCCATCCACATCCCTGCAATCCCTGCAATGCCAGTGGCCGCCCTCC 5447  
Qy 1912 ----- CCAGGGAACGTGCTGGCCACCCCTG 1935  
Db 5448 CCCAGAGGCCGTGTCCTGAATCGCTTCTGCCATCCAGGGAACGTGCTGGCCACCCCTG 5507  
Qy 1936 AATGGCAGTGTGCTGACAGCCCAAGCGAGGGCCCTGGGCCAGCTGCCCTCGGAG 1995  
Db 5508 AATGGAGTGTGCTGGACAGCCCAAGCGAGGGCCCTGGGCCAGCTGCCCTCGGAG 5567  
Qy 1996 GTCAGGGCCAGAAAGCTGCGAAACGTGCTGCTGAGCGCGGGGAGGGCTACATCGAC 2055  
Db 5568 GTCAGGGCCAGAAAGCTGCGAAACGTGCTGCTGAGCGCGGGGAGGGCTACATCGAC 5627  
Qy 2056 TTCCGCATT----- 2064  
Db 5628 TTCCGCATTGGTGAGCGGGGCCACAGGACAGGGCTGAGGTTGGCGCGGGGAGCCCTGG 5687  
Qy 2065 ----- GGAGACGGAGAGGACGACGAGACGGAGGGCGC 2099  
Db 5688 CGTCACTCTGCTGCTTTGCCCGCAGGAGACGGAGAGGACGACGAGACGGAGAGGGCGC 5747  
Qy 2100 AGGGACATGAGCCAGGTGAAGCCCGTGTCTCAAGGCAGAGCGCAGTCACATCATCGT 2159  
Db 5748 AGGGACATGAGCCAGGTGAAGCCCGTGTCTCAAGGCAGAGCGCAGTCACATCATCGT 5807  
Qy 2160 GTGGCAGGTGTCTACACCCCGCAGTGAAGCTGCTGCCCTGCCCTGGCCCGACCTGTACAT 2219  
Db 5808 GTGGCAGGTGTCTACACCCCGCAGTGAAGCTGCTGCCCTGCCCTGGCCCGACCTGTACAT 5867  
Qy 2220 AGGACCCCGACCACTGACCCCGCCCGCGCGGGGTAGCCAGCCAGCGCCCGCC 2279  
Db 5868 AGGACCCCGACCACTGACCCCGCCCGCGCGGGGTAGCCAGCCAGCGCCCGCC 5927  
Qy 2280 CCCTCTTCTAACCTCTCAACTGCACTTTTCACTGAGTCTGGCCCCCTCCAGCGGGCAGG 2339  
Db 5928 CCCTCTTCTAACCTCTCAACTGCACTTTTCACTGAGTCTGGCCCCCTCCAGCGGGCAGG 5987  
Qy 2340 GAGTCGGGGATGCGGATCACTGGGAGGAGGGAGGGGTGCTTCCACCCGAGGGGA 2399  
Db 5988 GAGTCGGGGATGCGGATCACTGGGAGGAGGGAGGGGAACCTTCCACCCGAGGGGA 6047  
Qy 2400 AGATGCTCTCGGACAGTTTCCCGGCGAGTCTTGCCAGTTCAGCCCGAGAGTCTCTCA 2459  
Db 6048 AGATGCTCTCGGACAGTTTCCCGGCGAGTCTTGCCAGTTCAGCCCGAGAGTCTCTCA 6107  
Qy 2460 AGTCCAGGSCACCTTGGGCCCCAGCGCAGGAGAAATCCGAGGTGGTCTTACCCCTG 2519  
Db 6108 AGTCCAGGSCACCTTGGGCCCCAGCGCAGGAGAAATCCGAGGTGGTCTTACCCCTG 6167  
Qy 2520 GGCCCTCTACTCCCCAGCACCCCTGGAGAGGCGAGGGCTCCCGCCGCGGAGGCTGCCT 2579  
Db 6168 GGCCCTCTACTCCCCAGCACCCCTGGAGAGGCGAGGGCTCCCGCCGCGGAGGCTGCCT 6227  
Qy 2580 GGCCCTGGGCCACCTCTGCATGCTGCTCATGGGGCCACCTGCCTCTCTGGGCCCTCACTC 2639  
Db 6228 GGCCCTGGGCCACCTCTGCATGCTGCTCATGGGGCCACCTGCCTCTCTGGGCCCTCACTC 6287  
Qy 2640 TGCCTAGGGAGCTGGGCCAGGCACTAGCCTTTGCCCAAGGAGGTGGGCTCAGGCTGCC 2699  
Db 6288 TGCCTAGGGAGCTGGGCCAGGCACTAGCCTTTGCCCAAGGAGGTGGGCTCAGGCTGCC 6347  
Qy 2700 CAGGTGCTGCACCCAGCGGCCCTTCTGTGGGGCCCTCCCGTCTCAAGCCCTCTATCT 2759  
Db 6348 CAGGTGCTGCACCCAGCGGCCCTTCTGTGGGGCCCTCCCGTCTCAAGCCCTCTATCT 6407  
Qy 2760 GTCTGTCCCCACCCAGTGTCCCTGCCAGGGAGCTGGGCATAAAGCAGGCGCCGG 2819  
Db 6408 GTCTGTCCCCACCCAGTGTCCCTGCCAGGGAGCTGGGCATAAAGCAGGCGCCGG 6467

Qy 2820 CTCCCTGGGGCAGCTGCTTGAGAACAGAGACTGCTACCCCCATCCTGCCCATGCAGGCAGG 2879  
Db 6468 CTCCCTGGGGCAGCTGCTTGAGAACAGAGACTGCTACCCCCATCCTGCCCATGCAGGCAGG 6527  
Qy 2880 CTCTTGCCAGCCCGTCTTGACCCGTGTCCCCCAGGCTCTGCTGGGCAGAAGACTCAC 2939  
Db 6528 CTCTTGCCAGCCCGTCTTGACCCGTGTCCCCCAGGCTCTGCTGGGCAGAAGACTCAC 6587  
Qy 2940 CTTGGAGGAGTGGGCCCTGGAGTCTCTCCCTCCAGAAAGCCCCCAGGGTGGATTCTC 2999  
Db 6588 CTTGGAGGAGTGGGCCCTGGAGTCTCTCCCTCCAGAAAGCCCCCAGGGTGGATTCTC 6647  
Qy 3000 AGSCTGCCAGGGCAGGCCCTCAGGCTCAGGAAGAGGGGAGGCCCTCGGCCTCTCCGGGAT 3059  
Db 6648 AGSCTGCCAGGGCAGGCCCTCAGGCTCAGGAAGAGGGGAGGCCCTCGGCCTCTCCGGGAT 6707  
Qy 3060 CAGTCTTAGGACACAGGCTCAGGCTCAGGTTGATGGGGATGATGTCTCCCGGGCCTG 3119  
Db 6708 CAGTCTTAGGACACAGGCTCAGGCTCAGGTTGATGGGGATGATGTCTCCCGGGCCTG 6767  
Qy 3120 CCTCTGCAACGGGCTCCACGAGGCCCTCAGTCCAGACACGCTACTAAGTGCCTAGGGTT 3179  
Db 6768 CCTCTGCAACGGGCTCCACGAGGCCCTCAGTCCAGACACGCTACTAAGTGCCTAGGGTT 6827  
Qy 3180 GCGCGCTGTGGCTGCTCCAGGGAGCAACAGAGAGGCCACCAAGCAGAGGCCCTGGGG 3239  
Db 6828 GCGCGCTGTGGCTGCTTCCAGGGAGCAACAGAGAGGCCACCAAGCAGAGGCCCTGGGG 6887  
Qy 3240 CTGAGGATGGAGCCGCCCTTATGAGGGCAGCCAGCCCGCAGAGGCGCAGACCCCTTGA 3299  
Db 6888 CTGAGGATGGAGCCGCCCTTATGAGGGCAGCCAGCCCGCAGAGGCGCAGACCCCTTGA 6947  
Qy 3300 CTGCTCTCCCTGCCCCAGTGGGCTCTCTGGCTATTCTTACCTTCCAGGCCACTGCAC 3359  
Db 6948 CTGCTCTCCCTGCCCCAGTGGGCTCTCTGGCTATTCTTACCTTCCAGGCCACTGCAC 7007  
Qy 3360 TCCTGTCTGGAGGCCCTTATGAGGGCAGCCAGCCCGCAGCCACCCCAACCCAGAGA 3419  
Db 7008 TCCTGTCTGGAGGCCCTTATGAGGGCAGCCAGCCCGCAGCCACCCCAACCCAGAGA 7067  
Qy 3420 AGCACAGATCTTGGGAGCTGCCCCACAAAGCCCGCTGGCCACCGAGGGCTGCAGCCGCT 3479  
Db 7068 AGCACAGATCTTGGGAGCTGCCCCACAAAGCCCGCTGGCCACCGAGGGCTGCAGCCGCT 7127  
Qy 3480 GCGCTGCGGCTTCTCCACACCCCTGCCACCTCCACTGTGATGTATGCTCGCTCCCTC 3539  
Db 7128 GCGCTGCGGCTTCTCCACACCCCTGCCACCTCCACTGTGATGTATGCTCGCTCCCTC 7187  
Qy 3540 GTCTGTCCCCCAGGATCTCGAAGTACTCCGGGCTAGCAGTGGGCGGCTGGGGAGG 3599  
Db 7188 GTCTGTCCCCCAGGATCTCGAAGTACTCCGGGCTAGCAGTGGGCGGCTGGGGAGG 7247  
Qy 3600 GGTGACGATCTCTCAGGCTTTGGCCCTGCAAGCAAAACCCACATATCTGTATGT 3659  
Db 7248 GGTGACGATCTCTCAGGCTTTGGCCCTGCAAGCAAAACCCACATATCTGTATGT 7307  
Qy 3660 AATAAATGTCTTAACGTCGTA 3680  
Db 7308 AATAAATGTCTTAACGTCGTA 7328

RESULT 13  
AF262046  
LOCUS AF262046 4545 bp mRNA linear ROD 12-DEC-2000  
DEFINITION Mus musculus sunday driver 2 (Syd2) mRNA, complete cds.  
ACCESSION AF262046  
VERSION AF262046.1 GI:11527194  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 4545)  
AUTHORS Bowman,A.B., Kamal,A., Ritchings,B.W., Philp,A.V., McGrail,M., Gindhart,J.G. and Goldstein,L.S.  
TITLE Kinesin-dependent axonal transport is mediated by the sunday driver (SYD) protein  
JOURNAL Cell 103 (4), 583-594 (2000)  
MEDLINE 20560743  
PUBMED 11106729  
REFERENCE 2 (bases 1 to 4545)  
AUTHORS Bowman,A.B., Philp,A.V., Ritchings,B.W., Kamal,A., McGrail,M., Gindhart,J.G. and Goldstein,L.S.B.  
TITLE Direct Submission  
JOURNAL Submitted (29-APR-2000) HHMI, University of California San Diego, 9500 Gilman Dr. m/c 0683, La Jolla, CA 92093-0683, USA  
FEATURES Location/Qualifiers  
source  
1. .4545  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL"  
/db\_xref="dbEST:AU051436"  
/db\_xref="taxon:10090"  
/clone="MNCb-2026"  
/sex="female"  
/tissue type="brain"  
/clone\_lib="Sugano mouse brain mncb"  
/dev\_stage="adult"  
/note="clone obtained from the Japanese Collection of Research Bioresources, National Institute of Infectious Disease, Division of Genetic Resources"  
1. .4545  
/gene="Syd2"  
110. .4120  
/gene="Syd2"  
/note="SYD2; related to Drosophila melanogaster sunday driver (SYD) protein"  
/codon\_start=1  
/product="sunday driver 2"  
/protein\_id="AAG36931.1"  
/db\_xref="GI:11527195"  
translation="MEIQMDEGGVVYQDDYCSGVMSERVSLAGSIYREFERLIH  
CYDEEVVKELMPLVNVLENLDSVLSENOHEHEVELEILLREDNEQLLTOYEREKALRKQ  
AEKFIIEEDALEQEKELQIQVEHYEFQTRQLELKAKYADQISRLEERESMKKEY  
NALHQHTEMIQTYVEHTERS KMQQVGGSGQTESLPGRSRKERPTSLNVFPLADGMV  
RAQMGKLVPA GDHWHLSDLGLQSSSYQC PNDEMSESGQSSAAATSTTGTSNTP  
TSSVPSAAVTPLNESLQPLGDYVSVTKNNKQARERNRNMVEQVTOEMRNVSIGMS  
SDEWSDVDIIDS TELPVCPELTRERTGSSPTQGI VNKAFGINTDSL YHELSTAGSE  
VIGDVEGADLLGEFSVRDDFFGMGKEVGNLLENSQLLETKNALNVKNDLIAKVDQ  
LSGEQVLEGELEAAKQAKVLENRILEEELKRVKSEAVTARPREVEEDVSSYL  
CTELDKIPMAQRRTFRVEMARVLMERNQYKERLMELQEA VRWTEMI RASRKHPSVQE  
KKKSTIQWFFSRLFS SSSPPPAKRSYPSVNIHYKSP TAAQFSQRRSHALCQISAGSR  
PLEFFPDDCTSSARREOKREYQVRHVRNDDGRLOACGWSLPAKYQLSPNGGQE  
DTRMKNVPVPVYCRPLVEKDPSTKLWCAAGVNL SGWKPHEDSSNGPKPVPGRDPLTC  
DREGEGEPKSTHPSPEKKAKETPEADATSSRVWILTSTLTTSKVIIIDANPGTIVD  
QFTVCNAHVL CISSIPAA SDDSDYPPGEMFLDSDVNPEDSGADGVLAGITLVGCATRCN  
VPRSNCSRGDTPVLDKGQGDVATTANGKVNPSQSTEEATEATEVDPGPSESEATTV  
RPGPLTEHVFTDPAPTPSSSTQPA SENGSESNGTIVQVPEPSGELSTTSSAAPT MW  
LGAQNGWLVVHSAVANWKKLHSIKLXDSVLSLVHKGRVLVALADGT LAIFHRGEDG  
QWDL SNVHLMDLGHPHHSIRCM AVNDRVWCGYKKNKVHVIQPKTMQIEKSFDAHPRE  
SQVRQLAWITGDGVWVSIRLDSLRLYHATHQLDQVDIIPYVSKMLGTGKLGFSFVR  
ITALLIAGNRLWVG TNGVVISIPLTETVVLHRGQLLGLRANKTSPTS GEGTRPGGII  
HVYGDSSDKAASSFIPYCSMAQAQLCFHGHRDAVKFFVSPGNVLATLNGSVLDSPS  
EGPGPAAPVADAEGQKLNALVLSGGEGYIDFRIGDGEDDETEECAGDVNQTKPSLSK  
DERTHIIIVWQVSYTPE"

gene  
CDS  
ORIGIN  
Query Match 41.8%; Score 1542.6; DB 10; Length 4545;  
Best Local Similarity 79.9%; Pred. No. 2.2e-216;  
Matches 1844; Conservative 0; Mismatches 459; Indels 6; Gaps 2;  
Qy 148 GATGGGTAGGAGCCAGGTTCTGTCGCCACGGCGCCTCCCTGCTCCCTGCAGCTGAGTCCC 207  
Db 2081 GACGGGAGGCTGCAGGCCCTGTGGTGGAGCCTGCCTGCCAAGTACAAGCAGCTGAGCCCC 2140

QY 208 AACGGGGCCAGGAGGACACGCGGATGAAGAACGTGCCGTGCCGTGCTACTGCCGCCCT 267  
Db 2141 AATGGAGGCCAGGAAGACACCCGGATGAAAAATGTCCTTCCCTGTGTACTGCCGCCCT 2200  
QY 268 CTGGTGGAGAGGACCCCAACCATGAAGCTGTGGTGTGCCGGGGCGTCAACCTGAGCGGG 327  
Db 2201 CTGGTGGAGAGGACCCCTTCGACAAAGCTGTGGTGTGTGCTGTGTTCAACTGAGTGG 2260  
QY 328 TGGAGGCCCAATGAGGACGACGCTGGGAATGGAGTCAAGCCAGCGCCAGCGCGATCCC 387  
Db 2261 TGGAGGCCCAATGAAGAGGACTCTAGCAATGGACCCCAAGCCTGTACCAGGTGAGACCT 2320  
QY 388 CTGACCTGGACCCGGAAGGAGACGGCGAGCCCAAGAGCGCCACACACGTCTCCCGAGAAG 447  
Db 2321 CTGACCTGTGACCGGGAAGGAGAGGGCAACCCCAAGAGCACACACCATCACCTGAGAAG 2380  
QY 448 AAGAAAGGCCAAGGAGCTCCCTGAAATGGACGCCACCTCCAGCGGGGTGTGGATCCTGACC 507  
Db 2381 AAGAAAGGCCAAGGAAACCCCTGAGGCAGATGTACCTCCAGTGGGTATGGATCCTCACC 2440  
QY 508 AGCACCTGACCAACAGCAAGTGGTGTATCATCGACGCCAACACCGCGGGCACGGTGGTG 567  
Db 2441 AGCACCTGACCAACAGCAAGTGGTGTATCATTTGATGCCAACCCAGCCAGGCACAATTGTG 2500  
QY 568 GACCAAGTTACCGTCTGCAACCGGCACGTCGTGTGTGTCATCTCCAGCATCCCCGGGCCAGC 627  
Db 2501 GATCAGTTACAGTCTGCAATGCCACGTCCTCTGTGTATCTCCAGCATTCCTGCGGCCAGT 2560  
QY 628 GACAGCGACTACCTCCCGGGGAGATGTTCTTGGACAGCGACGTGAACCCAGAGGACCCG 687  
Db 2561 GACAGTGACTATCCCCCTGGGGAGATGTTCTCTAGACAGTGTGAACCTGAAGATTCA 2620  
QY 688 GCGCAGATGGCGTGTCTGCCCGGTATCACCCCTGGTGGGTGTGCGCACCCCGCTGCAACGTG 747  
Db 2621 GGTGCTGATGGTGTGCTGGCTGGCATCACCCCTGGTGGGTGTGTACCCCGCTGCAATGTT 2680  
QY 748 CCGCGGAGCAACTGCTCTCTCCCGAGGGGACACCCCACTGCTAGACAAGGGGCAGGGGAG 807  
Db 2681 CCACGTAGCAACTGTTCTCTACGAGGAGACACCCCACTGCTGGAAGGGGCAGGGGGAT 2740  
QY 808 GTGCCACCATCGCCAA CGGGAAGGTCAAACCGTCCCACTCCAGAGGAGGCCACAGAG 867  
Db 2741 GTGGCGACCACTGCCAATGGGAAGGTCAAACCGTCCCACTCCAGAGGAAGCCACAGAA 2800  
QY 868 GCACAGGAGGTGCAGACCCCTGGGCCAGCGAGCCAGAGACAGCACATTCGGGCCCGGG 927  
Db 2801 GCCACGGAGGTGCAGACCCCTGGTCCAGCGAGTCAAGAACCGACAGTCCGGGCCCGGG 2860  
QY 928 CCTCTCAGAGACGCTCTTCACTGACCCAGCCCGGACCCCGTCTCTGGCCCCCAGCCT 987  
Db 2861 CCTCTCAGAGACGCTCTTCTTCTGACCCAGCACCCACCCCATCTCTCAGCACCCAGCCT 2920  
QY 988 GGCAGCGAGAACGGGCCAGAGCCCTGACAGCAGCAGACACCGCCAGAGCCAGAGCCAGC 1047  
Db 2921 GCCAGTGAGAAATGGGTCAAGTCCAAATGGCAACCATTTGTACAGCCTCAGGTGGAGCCCAGT 2980  
QY 1048 GGGAGCCCCACGGGAGCAGGACGAGTGTGCAACCCACCATGTGGCTGGGAGCCCCAGAAC 1107  
Db 2981 GGGGAACCTCTCAACAAACCAAGTAGCGCTGCACCCCACTATGTGGCTAGGAGCCCCAGAAT 3040  
QY 1108 GGCTGGCTCTATGTGCACTCGGCTGTGGCCAACTGGAAGAAGTGCCTGCACCTCCATCAAG 1167  
Db 3041 GGCTGGCTCTATGTGCATTCAGCGGTAGCCAACTGGAAGAAGTGTCTGCACCTCCATCAAG 3100  
QY 1168 CTGAAGGATTTCTGTGCTAGCCCTGGTGCATGTCAAAGCCCGTGTGCTGGTGGCTCTGGCG 1227  
Db 3101 CTAAAGACTCTGTGCTGAGCCCTGGTGCATGTCAAAGCCCGAGTGTGCTGGTAGCTCTTGCA 3160  
QY 1228 GACGGGACCCCTGGCCATCTTCCACCGTGGTGAAGATGGCCAGTGGGATCTGAGCAACTAT 1287  
Db 3161 GATGGGACCCCTGGCTATCTTCCATCTGTTGAGAGGATGGCCAGTGGGACCTGAGCAACTAC 3220  
QY 1288 CACCTAATGGACCTGGGCCACCCCGCACTCCATCCGCTGCATGGCTGTGTGTGTACGAC 1347

Db 3221 CACCTAATGGACCTGGGCCACCCACCACTCCATCCGTTGTCATGGCTGTGTGAATGAC 3280  
QY 1348 CGCGTGTGGTGTGGCTACAAGAACAAAGGTGCACGTGATCCAGCCCAAGACCATGACAGATA 1407  
Db 3281 CGAGTTTGGTGTGGCTACAAGAACAAAGGTGCATGTTATCCAGCCCAAGACAATGCAGATT 3340  
QY 1408 GAGAAGTCATTGACGCCCAACCCGCGCGGAGAGCCAGGTGCGGAGCTGGCGAGTGCCTGGATC 1467  
Db 3341 GAGAAATCATTTGATGCCCAACCAAGCGGGAAGCCAGGTACGTGAGCTGGCCTGGATC 3400  
QY 1468 GCGGATGGCGTATGGGTGTCATCCGCTGAGTCCACCCCTGAGGCTCTACCATGCACAC 1527  
Db 3401 GGTGATGGAGTGTGGTCTCTATTGCTTGGATTCTACCCCTTCGCGCTCTACCATGCTCAC 3460  
QY 1528 AGCACCAGCATCTACAGGACGTGGACATTGAGCCCTACGTACGCAAGATGCTAGGCACT 1587  
Db 3461 ACCCACCAGCACTGCAGGATGTGACATTGAGCCCTATGTTAGCAAGATGCTAGGAACC 3520  
QY 1588 GGCAAGCTGGTTTCTCCTTCGTACGCATCACGGCCCTGCTTGTGCGGGCAGCGCGCTC 1647  
Db 3521 GGCAAGCTGGGTTTCTCCTTCGTGCGCATCACAGCCTTACTCAFTGCAGGCAACCGTCTG 3580  
QY 1648 TGGGTGGGACCGGCAACCGAGTGGTCTATCTCCATCCCTGACAGAGACTGTGCTCCTG 1707  
Db 3581 TGGGTGGGCACTGGCAATGGGGTTGTCTCTCCATCCCTTGACTGAGACTGTGCTCCTG 3640  
QY 1708 CACGAGGCGAGCTCCTGGGGTCCGAGCCCAATGAAGACATCCCCACCTCTGGGGAGGGC 1767  
Db 3641 CATGAGGCGAGCTCCTAGGGTCCGAGCCCAAGACATCCCAACATCTGGGGAGGGG 3700  
QY 1768 GCGCGTCCGGGGCATCATCCACGTGTATGGGATGACAGCAGTGACAGGGCGGCCAGC 1827  
Db 3701 ACCCGCCAGGGGGCATCATCCATGTGTATGGGACGACAGCAGTGACAAAGCGCCAGT 3760  
QY 1828 AGCTTCATCCCTACTGCTCCATGGCCCGGAGGAGCTGCTTCCATGGGCACCGCGAT 1887  
Db 3761 AGTTTCATCCCTACTGCTCCATGGCAGCAGGCTCAGCTTTGCTTCCATGGGCACCGTGAT 3820  
QY 1888 GCCGTGAAGTCTTTGCTCGGTGCCAGGGAACGTGCTGGCCACCTGTAATGCGCAGTGTG 1947  
Db 3821 GCTGTCAAATCTTTGCTCTGTGCCAGGAATGTGCTGGCCACTCTCAATGGCAGTGTG 3880  
QY 1948 CTGGACAGCCAGCCGAGGGCCCTGGGCCAGCTGCCCTGCTCGGAGGTGCGAGGGCCAG 2007  
Db 3881 CTAGACAGCCCATCAGAGGGCCCTGGGCCCTGCTGCACCCGTTGCAGATGCTGAGGGCCAG 3940  
QY 2008 AAGCTGCGGAACGTGCTGCTGAGCGCGGGGAGGGCTACATCGACTTCCGATTGGA 2067  
Db 3941 AAGTTGAAGAAATGCACTGCTGCTGAGTGGTGAAGTTACATTGACTTCCGTATCGGA 4000  
QY 2068 GACGAGAGGACGACGAGACGAGGAGGGCGCAGGGGACATGAGCCAGGTGAAGCCCGCTG 2127  
Db 4001 GACGAGAGGATGATGAACATGAGGAATGTGCCGGGACGTGAACGACGACAAAGCCCTCG 4060  
QY 2128 CTGTCCAAGGCAGAGCGCAGTCACATCATCGTGTGGCAGGTGTCTTACACCCCGAGTGA 2187  
Db 4061 TTGTCCAAGGATGAGCGCAGCACATCATCGTGTGGCAGGTGTCTTACACCCCTGAGTG- 4119  
QY 2188 AGCTGTGCCCTGCTGCGCCCGACCTGTACATAGGACCCCGACCCACCTGACCCCGCCG 2247  
Db 4120 AGACCCCTGCTACCTGATGCCAACTGTACATAGGACCCCTACCTGCGCTGCCCTCCCGCCT 4179  
QY 2248 GGGCCGCGGGGTAGCCAGCAGGCGCGCGCCCTCTTCTAACCTCTCAACCTGCAGCT 2307  
Db 4180 GTTCCCTGGGCGAGCCA----GGTTCTGCTCCATCCCTTTTAACTCTCAACTGCACT 4234  
QY 2308 TTCACCTGAGTCTGGCCCTCCAGCGGCGAGGAGTGGGGGATGCGGATCAGCTGGGAG 2367  
Db 4235 TTGCTCTGAGGTCCAGCCCTAGCTGTAGAGAGGGATGAAGCAGGGTTTGGGGAGTGAG 4294  
QY 2368 GAGGAGGGGAGGGTGTCTTCCACCCGAGGGGAAGATGCTCTCGGGACAGTTTCCCGGGCA 2427

Db 4295 GCAGGAACACAGATGTCTTTCCCATCAGGGAGGAGAGAGCCTCTGAGACAATCTCTCCAGCA 4354  
QY 2428 GCTCCTGGCCAGCTTCCAGCCCGAGAGTCC 2456  
Db 4355 GCCCCTGGTCTTGACCCCAACATGAAGTCC 4383  
RESULT 14  
AB005662  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
CDS  
ORIGIN  
Query Match 41.5%; Score 1531.6; DB 10; Length 4173;  
Best Local Similarity 82.3%; Pred. No. 9.1e-215;

AB005662 4173 bp mRNA linear ROD 29-OCT-1999  
Mus musculus mRNA for JNK/SAPK-associated protein-1, complete cds.  
AB005662  
AB005662.1 GI:6141548  
JNK/SAPK-associated protein-1.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (sites)  
Ito, M., Yoshioka, K., Akechi, M., Yamashita, S., Takamatsu, N.,  
Sugiyama, K., Hibi, M., Nakabeppu, Y., Shiba, T. and Yamamoto, K. I.  
JSAPl, a novel jun N-terminal protein kinase (JNK)-binding protein  
that functions as a Scaffold factor in the JNK signaling pathway  
Mol. Cell. Biol. 19 (11), 7539-7548 (1999)  
99455010  
10523642  
2 (bases 1 to 4173)  
Yoshioka, K.  
Direct Submission  
Submitted (09-JUL-1997) Katsuji Yoshioka, Cancer Research  
Institute, Kanazawa University, Molecular Pathology; 13-1  
Takaramachi, Kanazawa, Ishikawa 920-0934, Japan  
{E-mail:katsuji@kenroku.ipc.kanazawa-u.ac.jp, Tel:81-76-265-2757,  
Fax:81-76-234-4517}  
Location/Qualifiers  
1. .4173  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="JSAPl"  
/tissue\_type="brain"  
107. .4024  
/codon\_start=1  
/product="JNK/SAPK-associated protein-1"  
/protein\_id="BAA85874.1"  
/db\_xref="GI:6141549"  
/translation="MMEIQMDEGGVVVYQDDYCSGVSMSERSVGLAGSIYREFRLI  
HCYDEEVVKELMPLVNVNLELDVLSNQHEVELELLREDNEQLLTQYERERKALRK  
QAEKFI EFEDALEQEKKELOIQVEHYEFQTRQLELKAKNYADQISRLEERSEMKKE  
YNALHORHTEMIQTYVEHIERSKMQQVGGSGQTESLPGRRKERPTSLNVFPLADGMC  
PNDEMSGQSSAAATPSITGTSNTPTSSVPSAAVPLNLSQLPLGDYVSVTKNNKQ  
AREKRNRMNEVQVTQEMRNVSIGMSDEWSDVDQDIIDSTPELDVCPETRLERTGSS  
PTQIVNKAFGINTDSLHELSTAGSEVIGDVDEGADLLGFSEVRDFFGMEKEVGNL  
LLENSQLLETKNALNVVKNDLIAKVDQLSGEQEVLKGELEAAKQAKVLENIKLEEE  
ELKRVKSEAVTARPREVEDVSSYLCTELDKIPMAQRRRFRVEMARVLMERNQYK  
ERLMELQEAVRWTEMRASRHPSPVQEKKSTIWQFFSRLFSSSSPPAKRSYPSVN  
IHYKSPTAAGFSQRRSHALCQISAGSRPLEFFPDDDCTSSARREKQREYQVREHVR  
NDDGRLQACGWSLPKAYKOLSPNGQOEDTRMKNVPVPCPLVEKDPSTKLWCAAGV  
NLGSKWPHEEDSSNGPKVPGRDPLTCDREGEPEPKSTHPSPEKKKAKETPEADATSS  
RWILTSLTTSKVVIDANQPTIVDQFTVCNAHLVCISSIPAAASDSDYPGEMFLD  
SDVNPEDSGADGVLGITLVGCATRCNVPRNSCSRGTDPVLDKGQGDVATTANGKVN  
PSQTEATEATEVDPGPSESEATTVPGLTEHVTDPAPTPSSSTQPSASENGSES  
NGTIVQVPEPSGELSTTSSAAPTMLGAQNGWLYVHSVANWKKLHSLKLDXSVL  
SLVHVKGRLVALADGTLAIFHRGEDGQWDLSNYHMLDLGHPHHS IRCMAVNDVRWC  
GYKNKHVQPKTMQIEKSFDAHPRRESQVRQLAWIGDVWVSIRLSDTLRLYHAHTH  
QHLQDVDIPIPVSKMLGTGKLGFSFVRITALLIAGNRLVGTGNGVVISIPLTETVVL  
HRGQLGLRANKTSPTSGETRPGGIHIVYGDSDSDKAASSFIPYCSMAQQLCFHGH  
RDAVKFFSVPGNVLATLNGSVLDSPEGPGPAAPAADAGQKLKNALVLSGEGYID  
FRIGDGEDDETECAGDVNQTKPSLSKAERSHIIWQVSYTPE"

| Matches 1785; Conservative 0; Mismatches 379; Indels 6; Gaps 2; |      |  |      |
|---|------|--|------|
| QY  | 148  | GATGGGTAGGAGCCAGGGTTCTGTCGCCACGGCGCCTCCCTGCTCCCTGCAGTGAGTCCC   | 207  |
| Db  | 1985 | GACGGGAGGCTGCAGGCCTGTGGGTGGAGCTGCCTGCCAAGTACAAGCAGCTGAGCCCC    | 2044 |
| QY  | 208  | AACGGGGCCAGGAGGACACGGGATGAAGAACGTGCCGGTGCCGGTGTACTGCCGCCCT     | 267  |
| Db  | 2045 | AATGGAGGCCAGGAAGACACCCGGATGAAAAATGTGCCTGTCCCTGTGTACTGTGCCCT    | 2104 |
| QY  | 268  | CTGGTGGAGAGGACCCCAACCATGAAGCTGTGGTGTCCCGGGCGTCAACCTGAGCGG      | 327  |
| Db  | 2105 | CTGGTGGAGAGGACCCCTTCGACAAAGCTGTGGTGTCTGTCTGTCTGTACTGTGAGTGG    | 2164 |
| QY  | 328  | TGGAGGCCCAATGAGGACGACGCTGGGAATGGAGTCAAGCCAGCGCCAGGCCGATCCC     | 387  |
| Db  | 2165 | TGGAAGCCACATGAAGAGGACTCTAGCAATGGACCCCAAGCCTGTACCAGGTGAGACCCCT  | 2224 |
| QY  | 388  | CTGACCTGCGACCGCGAAGGAGCGGCGAGCCCAAGAGCGCCACACACGTCTCCCGAGAAG   | 447  |
| Db  | 2225 | CTGACCTGTACCGGGAAGGAGAGCGGAACCCCAAGAGCACACACCCATCACCTGAGAAG    | 2284 |
| QY  | 448  | AAGAAGGCCAAGGAGCTCCCTGAATGGACGCGCACCTCCAGCCGGGTGTGATCTCTGACC   | 507  |
| Db  | 2285 | AAGAAGGCCAAGGAAAACCCCTGAGGCAGATGCTACCTCCAGTCGGGTATGATCCTCACC   | 2344 |
| QY  | 508  | AGCACCTTGACCAACAGCAAGTGTGTGATCATCGACGCCCAACAGCCGGGCAACGTGGTG   | 567  |
| Db  | 2345 | AGCACCTTGACCAACAGCAAGTGTGTGATCATTTGATGCCAACCCAGCCAGGCACAATTGTG | 2404 |
| QY  | 568  | GACCAGTTCACCGTCTGCAACGCGCACGTGCTGTGCATCTCCAGCATCCCGCGGCCAGC    | 627  |
| Db  | 2405 | GATCAGTTCACAGTCTGCAATGCCACGTCTCTGTATCTCCAGATTCTCTGGGCCAGT      | 2464 |
| QY  | 628  | GACAGCGACTACCTCCCGGGGAGATGTTCTTGGACGCGACGTGAACCCAGAGACCCCG     | 687  |
| Db  | 2465 | GACAGTGACTATCCCCCTGGGAGATGTTCTTAGACAGTGTGTGAACCCCTGAAGATTCA    | 2524 |
| QY  | 688  | GGCGCAGATGGCGTGTGGCCGCTATCACCTGTGGGCTGTGCCACCCGCTGCAACGTG      | 747  |
| Db  | 2525 | GGTGTGATGGTGTGCTGGCTGGCATCACCTGTGGGTGTGTACCCGCTGCAATGTT        | 2584 |
| QY  | 748  | CCGCGGAGCAACTGCTCCTCCCGAGGGGACACCCAGTGTAGACAAAGGGGCGAGGGGAG    | 807  |
| Db  | 2585 | CCACGTAGCAACTGTTCTCACGAGGAGACACCCCACTACTGGACAAAGGGGCGAGGGGAT   | 2644 |
| QY  | 808  | GTGGCCACCATCGCCCAACGGGAAGGTCAACCCGTCACAGTCCACAGAGGAGGCCACAGAG  | 867  |
| Db  | 2645 | GTGGCGACCACTGCCAATGGGAAGGTCAACCCGTCCTCAATCCACAGAAGAAGCCACAGAA  | 2704 |
| QY  | 868  | GCCACGAGGTGCCAGACCCTGGGCCCGAGCGAGCCAGACAGCCACATTGCGGCCCGGG     | 927  |
| Db  | 2705 | GCCACAGAGGTGCCAGACCCTGGTCCAGCGAGTCAAGAACACGACAGTCCGGCCCGGG     | 2764 |
| QY  | 928  | CCTCTCACAGAGCACGTCTTCACTGACCCAGCCCCCGACCCCGTCTCTGGCCCCCAGCCT   | 987  |
| Db  | 2765 | CCTCTCACAGAGCATGTCTTTACTGACCCAGCACCCACCCCATCTCTCCAGCACCCAGCCT  | 2824 |
| QY  | 988  | GGCAGCGAGAACCGGCCAGAGCCTGACACGACGACGACACGGCCAGAGCCAGGCCAGC     | 1047 |
| Db  | 2825 | GCCAGTGAGAAATGGGTCAGAATCCAATGGCACCATTTGTACAGCCTCAGTGGAGCCCAAT  | 2884 |
| QY  | 1048 | GGGGACCCACGGGAGCAGGACAGTGTGTGACCCACCATGTGGCTGGGAGCCCGAAGC      | 1107 |
| Db  | 2885 | GGGGAATCTCAACAACACCAAGTAGCGTGCACCCCACTATGTGGCTAGGAGCCCAAGAT    | 2944 |
| QY  | 1108 | GGCTGGCTCTATGTGCACTCGGCTGTGGCCAACTGGAGAAGTGCCTGCACTCCATCAAG    | 1167 |
| Db  | 2945 | GGCTGGCTCTATGTGCAATTGAGCGGTAGCCAACTGGAGAAGTGTCTGCACTCCATCAAG   | 3004 |
| QY  | 1168 | CTGAAGGATCTGTGCTGAGCCTGGTGATGTCAAAGGCGGTGTGTGGTGGCTCTGGGG      | 1227 |
| Db  | 3005 | CTAAAAGACTCTGTGCTGAGCCTGGTGCATGTCAAAGGCGGAGTGTGTGTAGTCTTTGCA   | 3064 |

|    |      |  |      |
|----|------|--|------|
| QY | 1228 | GACGGGACCCTGGCCATCTTCCACCGTGTGAAGATGGCCAGTGGATCTGAGCAACTAT     | 1287 |
| Db | 3065 | GATGGGACCCTGGCTATCTTCCATCGTGGAGAGGATGGCCAGTGGACCTGAGCAACTAC    | 3124 |
| QY | 1288 | CACCTAATGACCTGGGCCACCCCGCACCACTCCATCCGCTGCATGGCTGTGTGTACGAC    | 1347 |
| Db | 3125 | CACCTAATGACCTGGGCCACCCACACCACTCCATCCGCTGCATGGCTGTGTGTGAATGAC   | 3184 |
| QY | 1348 | CGCGTGTGTGTGTGTACAAAGAACAGGTGCACGTCAATCCAGCCCAAGACCAATGCAGATA  | 1407 |
| Db | 3185 | CGAGTTTGTGTGTGTGTACAAAGAACAGGTGTGTATCCAGCCCAAGACCAATGCAGATT    | 3244 |
| QY | 1408 | GAGAAAGTCATTTGACGCCACCGCGCGGGAGAGCCAGGTGCGGCAGCTGGCGTGGATC     | 1467 |
| Db | 3245 | GAGAAATCATTTGATGCCACCCCAAGCGGGAAGCCAGGTACGTGAGTGGCTGGATC       | 3304 |
| QY | 1468 | GGCGATGGCGTATGGGTGTCCATCCGCTTGGACTCCACCTGAGGCTCTACCATGCACAC    | 1527 |
| Db | 3305 | GGTGATGGAGTGTGGGTCTCTATTCGTTGGATTCTACCCCTCGGCTCTACCATGCTCAC    | 3364 |
| QY | 1528 | AGCCACCAGCATCTACAGGACGTGGACATTTAGCCCTACGTACAGCAAGATGCTTAGGCACT | 1587 |
| Db | 3365 | ACCACCAGCACTGCAGGATGTGGACATTTAGCCCTATGTTAGCAAGATGCTTAGGAACC    | 3424 |
| QY | 1588 | GGCAAGCTGGGTTCTCTTCGPACGCATCACGGCCCTGTCTTTCGCGGGCGAGCGGGCTC    | 1647 |
| Db | 3425 | GGCAAGCTGGGTTCTCTTCGTGCGCATCACAGCCTTACTCAATTGCAGGCAACCGTCTG    | 3484 |
| QY | 1648 | TGGTGGGCACCGGCAACGGAGTGGTTCATCTCCATCCCCCTGACAGAGACTGTGGTCTG    | 1707 |
| Db | 3485 | TGGTGGGCACCTGGCAATGGGGTGTCTATCTCCATCCCCCTGACTGAGACTGTGGTCTG    | 3544 |
| QY | 1708 | CACCGAGGCCAGCTCCTGGGGTCCGAGGCCAATAAGACATCCCCACCTCTGGGAGGGC     | 1767 |
| Db | 3545 | CATCGAGGCCAGCTCCTAGGGTCCGAGGCCAACAAGACATCCCCAATCTGGGAGGGG      | 3604 |
| QY | 1768 | GCCGCTCCCGGGGCATCATCCAGTGTATGGCGATGACAGCAGTGACAGGGCGGCAGC      | 1827 |
| Db | 3605 | ACCGGCCAGGGGCATCATCCATGTGTATGGGACGACAGCAGTGACAAGGCCGCCAGT      | 3664 |
| QY | 1828 | AGCTTTCATCCCTTACTGCTCCATGGCCAGGCCAGCTATGCTTCCATGGGCACCCGAT     | 1887 |
| Db | 3665 | AGTTTTCATCCCTTACTGCTCCATGGCACAGGCTCAGCTTTGCTTCCATGGGCACCCGAT   | 3724 |
| QY | 1888 | GCCGTGAAGTTCCTTGTCTCGGTGCCAGGGAACGTGCTGGCCACCTGAATGGCAATGTG    | 1947 |
| Db | 3725 | GCTGTCAAATTCCTTGTCTCTGTGCCAGGAAATGTGTGGCCACTCTCAATGGCAATGTG    | 3784 |
| QY | 1948 | CTGGAAGCCAGCCAGGGCCCTGGGCCAGCTGCCCTGCTCCGAGGTTCGAGGGCCAG       | 2007 |
| Db | 3785 | CTAGACAGCCCATCAGAGGGCCCTGGGCCCTGCTGCAACCCGCTGCAGATGCTGAGGGCCAG | 3844 |
| QY | 2008 | AAGCTGCGGAACGTGCTGGTGTGCTGAGCGGGGAGGGCTACATCGACTTCCGCAATTGGA   | 2067 |
| Db | 3845 | AAGTTGAAGAATGCACTGGTGTGCTGAGTGGTGAAGGTACATTGACTTCCGTATCGGA     | 3904 |
| QY | 2068 | GACGAGAGGACGACGAGACGGAGGGCGGCGAGGACATGAGCCAGGTGAAGCCCGTG       | 2127 |
| Db | 3905 | GACGAGAGGATGATGAAACTGAGGAATGTGCCGGGACGTGAACACAGACAAAAGCCCTCG   | 3964 |
| QY | 2128 | CTGTCCAAGGCAGAGCGCAGTCACATCATCTGTGGCAGGTGTCTTACACCCCGAGTGA     | 2187 |
| Db | 3965 | TTGTCCAAGGCTGAGCGCAGCCACATCATCTGTGGCAGGTGTCTTACACCCCTGAGTG-    | 4023 |
| QY | 2188 | AGCTGCTGCCCTGCTGGCCCGACCTGTACATAGGACCCCGACCACTGACCCCGGCC       | 2247 |
| Db | 4024 | AGACCTGTCTTACCTGATGCCAACTGTACATAGGACCTTACCTGCTGCTCCCGCCT       | 4083 |
| QY | 2248 | GGCCCGCGGGGTAGCCAGCCAGGGCGCGCCCGCTCTTCTAACTCTCAACCTGCACT       | 2307 |
| Db | 4084 | GTTCCCTGGGCGAGCCA----GGTTCGTCCATCCCCCTTTTAACTCTCAACTTGCAGCT    | 4138 |

QY 2308 TTCACCTGAG 2317  
 ||| |||||  
 Db 4139 TTTGCCTGAG 4148

RESULT 15  
 AB043123  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

AB043123 3945 bp mRNA linear ROD 13-OCT-2000  
 Mus musculus Jsap1 mRNA for JNK/SAPK-associated protein 1b,  
 complete cds.  
 AB043123  
 AB043123.1 GI:10801122  
 .  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (sites)  
 Ito,M., Akechi,M., Hirose,R., Ichimura,M., Takamatsu,N., Xu,P.,  
 Nakabeppu,Y., Tadayoshi,S., Yamamoto,K. and Yoshioka,K.  
 Isoforms of JSAP1 scaffold protein generated through alternative  
 splicing  
 Gene 255 (2), 229-234 (2000)  
 MEDLINE 20480689  
 PUBMED 11024282  
 2 (bases 1 to 3945)  
 Yoshioka,K.  
 Direct Submission  
 Submitted (22-MAY-2000) Katsuji Yoshioka, Cancer Research  
 Institute, Kanazawa University, Molecular Pathology; 13-1  
 Takaramachi, Kanazawa, Ishikawa 920-0934, Japan  
 (E-mail:katsuji@kenroku.kanazawa-u.ac.jp, Tel:81-76-265-2757,  
 Fax:81-76-234-4517)

FEATURES  
 source Location/Qualifiers  
 1..3945  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /tissue\_type="brain"  
 /clone\_lib="mouse brain cDNA library"  
 1..3945  
 /gene="Jsap1"  
 1..3945  
 /gene="Jsap1"  
 /codon\_start=1  
 /product="JNK/SAPK-associated protein 1b"  
 /protein\_id="BAB16674.1"  
 /db\_xref="GI:10801123"  
 /translation="MMEIQMDEGGVVVYDDYC SGVSMSERVSLAGSIYREFERLI  
 HCYDEEVKELMPLVNVLENLDSVLSENQEHEVELELLREDNEQLLTQYEREKALRK  
 QAEEKFIEFDALAEQKKELQI QVEHYEFTRQLELKAKNYADQISRLREERSEMKKE  
 YNALHQRHTEMIQTYYEHILERSKMQQVGSGGTLESSLPGRSPQRSWRKSRRKERPTSLN  
 VFPLADGMCPDENMESGQSAAATPSTTGTKSNTPTSVPSSAATPLNESLQPIGDY  
 VSVTKNNKAQREKNRNMVEQVTQMERNVSI GMSSDSDVDQIIDSTPELDVCPE  
 TRLTGSGSTQGIVNKAFGINTDSLXHELSTAGSEVICDVDEGADLIGFEFSVRDDFF  
 GMGKEGVNLLLENSQLLETKNALNVVKNDLIAKDQLSGEQEVLKGELEAAQAOKVKL  
 ENRIKELEBELKRKVSEA VTARPREVEDVSYSLCTELDKI PMAQRRTFRVEMAR  
 VLMRNQYKERLMELQEA VRWTEMIRASRHPSVQEKKSTIWFFSRFLPSSSSPPP  
 AKRSYPVNIHYKSPTAAGFSQRRSHALCQISAGSRPLEFFPD DDC TS SARREQKRE  
 YRQREHRNRNDGRLQACGWSLP AKYKQLSPNGGQEDTRMKVTPVFPVCRPLVEKDPS  
 TKLWCAGVNLSGWKPHEDSSNGPKPVPRDPLTCDRGEGETKSTHPSPPEKKAKE  
 TP EADATSSRVWILTSTLTSTSKVIIIDANQP GTIVDQFTVCNAHL CISSI PAASDS  
 YPPGEMFLDSVNPEDSGADVLAGITLVGCATRCNVRPNSCSSRGDT PVLDKGQGDY  
 ATTANGKDPSPSTEATEATEVPDPGPSEATTVRPGPLTEHVFTDPA PT PSSSTQ  
 PASENGSENGTI VQPQEPFSGELSTTSSAAP TMWLGAQNGLYVHS AVANWKCLH  
 SIKLKDSLVL SVHKGRVLVALADGT LAIFHRGEDGQWDL SNYHLM DLGH PHHS IRCM  
 AVNDVMCGYKHNKHVIQPTMQIEKSFDAHPRESQVRLAWIGDWVWVSIRLDST  
 LRLYHAHTHQKLQDV DIEPVVSKMLGTGLGF SFVRTALLIAGNRLWVGTVGNVVIS  
 IPLTETVVLHRGQLLG LRANKTSPTSGEGTRPGGI IHVYGDDSSDKAASSFI PYCSMA  
 QAOLCFHGHRDAVKFFVSPGNVLATLNGSVLDSPSEGPGPAAPA DAEGQKLKNALV  
 LSGGEGYIDFRIGDGEDDETCEACGVDNQT KP SL SAERSHII IVWQVSYTPE"

gene  
 CDS

ORIGIN

|    |      |  |      |
|----|------|--|------|
| Db | 2926 | CTAAAAGACTCTGTGCTGAGCCTGGTGCATCTCAAAGGCCGAGTCTGGTAGCTCTTGCA      | 2985 |
| Qy | 1228 | GACGGGACCCCTGGCCATCTTCCACCGTGGTGAAGATGCCCACTGGGATCTGAGCAACTAT    | 1287 |
| Db | 2986 | GATGGGACCCCTGGCTATCTTCCATCGTGGAGAGGATGCCCACTGGGACCTGAGCAACTAC    | 3045 |
| Qy | 1288 | CACCTAATGGAACCTGGCCACCCGCCACCACTCCATCCGCTGCATGGCTGTTGTGTACGAC    | 1347 |
| Db | 3046 | CACCTAATGGAACCTGGCCACCCACACCACTCCATCCGCTGCATGGCTGTTGTGAATGAC     | 3105 |
| Qy | 1348 | CGCGTGTGGTGTGGCTACAAGAACAAAGGTGCAGTCAATCCAGCCCCAAGACCATGCAGATA   | 1407 |
| Db | 3106 | CGAGTTTGGTGTGGCTACAAGAACAAAGGTGCATGTTATCCAGCCCCAAGACAATGCAGATT   | 3165 |
| Qy | 1408 | GAGAAAGTCATTTGACGCCCCACCCGCCGCGGGAGAGCCAGGTGGCGCAGCTGGCGTGGATC   | 1467 |
| Db | 3166 | GAGAAATCATTTGATGCCCCACCCAAAGCGGGAAAGCCAGGTAGTACGTGAGCTGGCCTGGATC | 3225 |
| Qy | 1468 | GGCGATGGCGGTATGGGTGTCCATCCGCCCTGGACTCCACCCTGAGGCTCTACCATGCACAC   | 1527 |
| Db | 3226 | GGTGATGGAGTGTGGGTCTCTATTCGCTTGGATTCTACCCCTCGGCTCTACCATGCTCAC     | 3285 |
| Qy | 1528 | ACGCACCAGCATCTACAGGACGTGGACATTGAGCCCTACGTACAGCAAGATGCTAGGCACT    | 1587 |
| Db | 3286 | ACCCACCAGCACCTGCAGGATGTGGACATTGAGCCCTATGTTAGCAAGATGCTAGGAACC     | 3345 |
| Qy | 1588 | GGCAAGCTGGGTTTCTCCTTCGTAGCATCACGGCCCTGCTTGTTCGGGGCAGCCGGCTC      | 1647 |
| Db | 3346 | GGCAAGCTGGGCTTCTCCTTCGTGGCATCACAGCCTTACTCATTCAGGCAACCGTCTG       | 3405 |
| Qy | 1648 | TGGGTGGGCAACGGCAACGGAGTGGTCACTCTCCATCCCCCTGACAGAGACTGTGTCCTG     | 1707 |
| Db | 3406 | TGGGTGGGCACTGGCAATGGGGTGTGTCACTCCATCCCCCTTGACTGAGACTGTGTCCTG     | 3465 |
| Qy | 1708 | CACCGAGGCCAGCTCCTGGGGCTCCGAGGCCAATAAGACATCCCCCACCTCTGGGGAGGGC    | 1767 |
| Db | 3466 | CATCGAGGCCAGCTCCTAGGGCTCCGAGGCCAACAGACATCCCCAACATCTGGGGAGGGG     | 3525 |
| Qy | 1768 | GCCCGTCCCGGGGCATCATCCAGTGTATGGCGATGACAGCAGTACAGGGCGGCCAGC        | 1827 |
| Db | 3526 | ACCCGCCAGGGGGCATCATCCATGTGTATGGGGACGACAGCAGTACAGGGCGGCCAGT       | 3585 |
| Qy | 1828 | AGTTTCATCCCCCTACTGCTCCATGGCCCCAGGCCAGCTATGCTTCCATGGGCACCCCGAT    | 1887 |
| Db | 3586 | AGTTTCATCCCCCTACTGCTCCATGGCACAGGCTCAGCTTTGCTTCCATGGGCACCCGTAT    | 3645 |
| Qy | 1888 | GCCGTGAAGTTCTTTGTCTCGGTGCCAGGGAACTGTCTGGCCACCTGAAATGGCAGTGTG     | 1947 |
| Db | 3646 | GCTGTCAAAATTCCTTTGTCTCTGTGCCAGGAAATGTGTGGCCACTCTCAATGGCAGTGTG    | 3705 |
| Qy | 1948 | CTGGACAGCCCAAGCCGAGGGCCCTGGGCCAGCTGCCCTGCCTCGGAGGTCGAGGGCCAG     | 2007 |
| Db | 3706 | CTAGACAGCCCATCAGAGGGCCCTGGGCCCTGCTGCACCCGCTGCAGATGCTGAGGGCCAG    | 3765 |
| Qy | 2008 | AAGTCGCGAACTGTCTGTGTGTGTCGAGCGCGGGAGGGCTACATCGACTTCCGCAATTGA     | 2067 |
| Db | 3766 | AAGTTGAAGAAATGCACTGGTGTGTGAGTGGTGAAGGTTACATTGACTTCCGTATCGGA      | 3825 |
| Qy | 2068 | GACGGAGAGGACGACAGACGGAGGGCGCAGGGGACATGAGCCAGGTGAAGCCCGTG         | 2127 |
| Db | 3826 | GACGGAGAGGATGATGAAACTGAGGAATGTGCCGGGACGTGAACCAAGAAAGCCCTCG       | 3885 |
| Qy | 2128 | CTGTCCAAGGCAGAGCGCAGTCACTCATCGTGTGGAGGTGTCTTACACCCCGAGTGA        | 2187 |
| Db | 3886 | TTGTCCAAGGCTGAGCGCAGCCACATCATCGTGTGGCAGGTGTCTTACACCCCTGAGTGA     | 3945 |

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 11:03:47 ; Search time 59 Seconds  
(without alignments)  
3117.600 Million cell updates/sec

Title: US-10-019-495-9  
Perfect score: 3473  
Sequence: 1 MKNVPVPCRLVEKDPTM.....VLSKAERSHIIVQVSYTPE 651

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID      | Description         |
|------------|--------|-------------|--------|------------|---------------------|
| 1          | 3473   | 100.0       | 651    | 3 AAB29652 | Aab29652 Human mem  |
| 2          | 3468   | 99.9        | 1066   | 4 AAM41910 | Aam41910 Human pol  |
| 3          | 3468   | 99.9        | 1069   | 3 AAB42717 | Aab42717 Human ORF  |
| 4          | 3465   | 99.8        | 1014   | 4 AAM40124 | Aam40124 Human pol  |
| 5          | 3369   | 97.0        | 1069   | 4 ABG09904 | Abg09904 Novel hum  |
| 6          | 3191   | 91.9        | 1305   | 3 AAB12875 | Aab12875 Murine JN  |
| 7          | 3191   | 91.9        | 1314   | 3 AAB12876 | Aab12876 Murine JN  |
| 8          | 3191   | 91.9        | 1336   | 3 AAB12878 | Aab12878 Murine JN  |
| 9          | 3191   | 91.9        | 1337   | 3 AAB12877 | Aab12877 Murine JN  |
| 10         | 2757.5 | 79.4        | 809    | 4 ABG09902 | Abg09902 Novel hum  |
| 11         | 1821.5 | 52.4        | 1307   | 6 ADA45192 | Ada45192 Murine JLP |
| 12         | 1814.5 | 52.2        | 1311   | 6 ADA45190 | Ada45190 Human JLP  |
| 13         | 1732.5 | 49.9        | 1277   | 4 ABB11784 | Abb11784 Human sec  |
| 14         | 1467   | 42.2        | 1165   | 6 ADA45195 | Ada45195 Murine JL  |
| 15         | 1446.5 | 41.6        | 828    | 2 AAW64473 | Aaw64473 Human sec  |
| 16         | 1446.5 | 41.6        | 828    | 4 AAB90749 | Aab90749 Human EC1  |
| 17         | 1152   | 33.2        | 1195   | 4 ABB63694 | Abb63694 Drosophil  |
| 18         | 943    | 27.2        | 734    | 4 AAB95634 | Aab95634 Human pro  |
| 19         | 807.5  | 23.3        | 235    | 3 AAB54174 | Aab54174 Human pan  |
| 20         | 675.5  | 19.5        | 1000   | 6 ADA45196 | Ada45196 Murine JL  |
| 21         | 623.5  | 18.0        | 200    | 4 AAB94656 | Aab94656 Human pro  |
| 22         | 511    | 14.7        | 118    | 5 ABB89836 | Abb89836 Human pol  |
| 23         | 462    | 13.3        | 260    | 4 ABG09899 | Abg09899 Novel hum  |
| 24         | 462.   | 13.3        | 592    | 4 ABG15607 | Abg15607 Novel hum  |
| 25         | 460.5  | 13.3        | 766    | 7 ADB79917 | Adb79917 Human put  |

|    |       |      |      |            |                    |
|----|-------|------|------|------------|--------------------|
| 26 | 442.5 | 12.7 | 1935 | 6 ABJ19396 | Abj19396 Human int |
| 27 | 442.5 | 12.7 | 2063 | 5 ABB90736 | Abb90736 Human tum |
| 28 | 442.5 | 12.7 | 2063 | 6 ABU54443 | Abu54443 Human tum |
| 29 | 436.5 | 12.6 | 1101 | 4 ABB63244 | Abb63244 Drosophil |
| 30 | 333.5 | 9.6  | 1195 | 6 ABU11484 | Abu11484 Human MDD |
| 31 | 333.5 | 9.6  | 1233 | 5 ABP55147 | Abp55147 Neurodeve |
| 32 | 333.5 | 9.6  | 1240 | 6 ABU54572 | Abu54572 Human NOV |
| 33 | 306   | 8.8  | 988  | 4 ABG08836 | Abg08836 Novel hum |
| 34 | 306   | 8.8  | 988  | 7 ADE09396 | Ade09396 Novel pro |
| 35 | 305.5 | 8.8  | 1129 | 4 ABG08838 | Abg08838 Novel hum |
| 36 | 265   | 7.6  | 1783 | 4 ABB63930 | Abb63930 Drosophil |
| 37 | 233   | 6.7  | 596  | 4 AAB92909 | Aab92909 Human pro |
| 38 | 191.5 | 5.5  | 91   | 4 AAU32873 | Aau32873 Novel hum |
| 39 | 186.5 | 5.4  | 119  | 5 ABB89345 | Abb89345 Human pol |
| 40 | 185.5 | 5.3  | 52   | 4 ABB40157 | Abb40157 Peptide # |
| 41 | 185.5 | 5.3  | 52   | 4 AAM33802 | Aam33802 Peptide # |
| 42 | 185.5 | 5.3  | 52   | 4 AAM73607 | Aam73607 Human bon |
| 43 | 185.5 | 5.3  | 52   | 4 AAM60922 | Aam60922 Human bra |
| 44 | 185.5 | 5.3  | 52   | 4 ABG55340 | Abg55340 Human liv |
| 45 | 185.5 | 5.3  | 52   | 5 ABG43477 | Abg43477 Human pep |

ALIGNMENTS

RESULT 1  
AAB29652  
ID AAB29652 standard; protein; 651 AA.  
XX  
AC AAB29652;  
XX  
DT 23-FEB-2001 (first entry)  
XX  
DE Human membrane-associated protein HUMAP-9.  
XX  
KW Human membrane-associated protein; HUMAP; transgenic organism;  
KW drug screening; cell signalling modulator; agonist; antagonist;  
KW cell differentiation modulator; cell proliferation modulator;  
KW cell proliferative disorder; cancer; cell differentiation disorder;  
KW developmental disorder; cell signalling disorders; endocrine disorder;  
KW hyperpituitarism; hypothyroidism; hyperparathyroidism; infection;  
KW pancreatic disorder; diabetes mellitus; immunological disorder;  
KW hereditary neuropathy; gonadal steroid hormone associated disorder;  
KW infertility.  
XX  
OS Homo sapiens.  
XX  
PN WO200065054-A2.  
XX  
PD 02-NOV-2000.  
XX  
PF 20-APR-2000; 2000WO-US010884.  
XX  
PR 23-APR-1999; 99US-0130694P.  
PR 23-JUN-1999; 99US-0140580P.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
PI Hillman JL, Bandman O, Tang YT, Lal P, Yue H, Reddy R;  
PI Azimzai Y, Baughn MR;  
XX  
DR WPI; 2000-687346/67.  
DR N-PSDB; AAC64282.  
XX  
PT Human membrane-associated protein, useful for diagnosis and treatment of  
PT cell signaling, cell differentiation and cell proliferation disorders  
PT such as cancer, and for identifying agonists and antagonists.  
XX  
PS Claim 1; Page 78-79; 99pp; English.  
XX  
CC The invention relates to 17 human membrane-associated proteins, HUMAP-1  
CC to HUMAP-17 (AAB29644-B29660) and the cDNAs encoding them (AAC64274-  
CC C64290). The invention also relates to expression constructs, host cells

CC and transgenic organisms comprising a HUMAP nucleic acid sequence; the  
CC recombinant preparation of a HUMAP; methods of screening compounds for  
CC their ability to modulate HUMAP activity or expression; and  
CC pharmaceutical compositions comprising a HUMAP protein, a HUMAP agonist  
CC or HUMAP antagonist. The HUMAPs acts as modulators of cell signalling,  
CC differentiation and proliferation. A HUMAP is useful for screening a  
CC compound for effectiveness as an agonist or antagonist of HUMAP activity.  
CC The protein, or the identified agonist or antagonist is useful for  
CC treating a disease or condition associated with decreased or increased  
CC expression of functional HUMAP. A HUMAP nucleic acid is useful for  
CC screening a compound for its ability to alter expression of that  
CC particular HUMAP gene. A wide variety of disease may be treated using  
CC compositions of the invention. These diseases include cell proliferative  
CC disorders (e.g., actinic keratosis, arteriosclerosis); cancer (e.g.,  
CC breast, bladder, bone marrow, brain and uterine cancer); cell  
CC differentiation disorders, in particular developmental disorders (e.g.,  
CC renal tubular acidosis, anaemia, Cushing's syndrome, achondroplasia,  
CC epilepsy, and muscular dystrophy); cell signalling disorders, in  
CC particular endocrine disorders such as hypothalamus and pituitary  
CC disorders resulting from lesions such as thrombosis; disorders associated  
CC with hyperpituitarism (e.g., acromegaly); disorders associated with  
CC hypothyroidism (e.g., goitre); hyperparathyroidism; pancreatic disorders  
CC such as type I or type II diabetes mellitus; infections; immunological  
CC disorders; hereditary neuropathies (e.g., neurofibromatosis); and  
CC disorders associated with gonadal steroid hormones (e.g., infertility,  
CC endometriosis, polycystic ovary syndrome, osteoporosis, Leydig cell  
CC deficiency and gynecomastia). Antibodies which specifically bind HUMAP  
CC may be used for the diagnosis of disorders associated with the expression  
CC of HUMAP, or in assays to monitor patients being treated with HUMAP or  
CC agonists, antagonists or inhibitors of HUMAP. The present sequence  
CC represents a HUMAP of the invention

XX Sequence 651 AA;

Query Match 100.0%; Score 3473; DB 3; Length 651;  
Best Local Similarity 100.0%; Pred. No. 4.1e-270;  
Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKNVPVPVYCRPLVEKDPTMKLWCAAGVNLGWRPNEDDAGNGVKPAPGRDPLTCDEGD 60  
DB 1 MKNVPVPVYCRPLVEKDPTMKLWCAAGVNLGWRPNEDDAGNGVKPAPGRDPLTCDEGD 60  
QY 61 GEPKSAHTSPEKKAKELPEMDATSSRVWILTSTLTTSKWIIDANQPGTVVDQFTVCNA 120  
DB 61 GEPKSAHTSPEKKAKELPEMDATSSRVWILTSTLTTSKWIIDANQPGTVVDQFTVCNA 120  
QY 121 HVLCTISSIPAASDSYPPGEMFLDSVNPEDPGADGVLGITLVGCATRCNVPRNSCSR 180  
DB 121 HVLCTISSIPAASDSYPPGEMFLDSVNPEDPGADGVLGITLVGCATRCNVPRNSCSR 180  
QY 181 GDTPLVDKGOGEVATIANGKNPQSQTETEATEVEPDGPPSEPETATLRPGPLTEHVFT 240  
DB 181 GDTPLVDKGOGEVATIANGKNPQSQTETEATEVEPDGPPSEPETATLRPGPLTEHVFT 240  
QY 241 DPAPTSSGPQPGSENGPEPDSSSTRPEPEPSGDPTGAGSSAAPTMWLGQNGWLYVHSA 300  
DB 241 DPAPTSSGPQPGSENGPEPDSSSTRPEPEPSGDPTGAGSSAAPTMWLGQNGWLYVHSA 300  
QY 301 VANWKKCLHSIKLSDVLSLVHVKGRLVALADGTTAIFHRGEDGQWDLSNYHLMDLGHP 360  
DB 301 VANWKKCLHSIKLSDVLSLVHVKGRLVALADGTTAIFHRGEDGQWDLSNYHLMDLGHP 360  
QY 361 HHSIRCMVAVYDRVWVCYKKNKHVIOQKTMQIEKSFDAHPRESQVQLAWIGDGVWVSI 420  
DB 361 HHSIRCMVAVYDRVWVCYKKNKHVIOQKTMQIEKSFDAHPRESQVQLAWIGDGVWVSI 420  
QY 421 RLDSTLRLYHAHTHQHLDQVDIEPYVSKMLGTGKLGFSFVRITALLVAGSRLWVGTGNGV 480  
DB 421 RLDSTLRLYHAHTHQHLDQVDIEPYVSKMLGTGKLGFSFVRITALLVAGSRLWVGTGNGV 480  
QY 481 VISIPLTETVVLHVGQLGLRANKTSPTSGEGARPGGIHVYDGDSSDRAASSFIPYCSM 540  
DB 481 VISIPLTETVVLHVGQLGLRANKTSPTSGEGARPGGIHVYDGDSSDRAASSFIPYCSM 540

QY 541 AQAQLCFHGHRDAVKFFVSVPCNVLATLNGSVLDSPAEGPGPAAPASEVEGQKLRNVLVL 600  
DB 541 AQAQLCFHGHRDAVKFFVSVPCNVLATLNGSVLDSPAEGPGPAAPASEVEGQKLRNVLVL 600  
QY 601 SGGEGYIDFRIGDGEDDETEEGAGDMSQVKPVLskaershiwqvSYTPE 651  
DB 601 SGGEGYIDFRIGDGEDDETEEGAGDMSQVKPVLskaershiwqvSYTPE 651

RESULT 2

AAM41910  
ID AAM41910 standard; protein; 1066 AA.

XX AAM41910;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 6841.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US034263.

XX 23-DEC-1999; 99US-00471275.

PR 21-JAN-2000; 2000US-00488725.

PR 25-APR-2000; 2000US-00552317.

PR 20-JUN-2000; 2000US-00598042.

PR 19-JUL-2000; 2000US-00620312.

PR 03-AUG-2000; 2000US-00653450.

PR 14-SEP-2000; 2000US-00662191.

PR 19-OCT-2000; 2000US-00693036.

PR 29-NOV-2000; 2000US-00727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
PI Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

DR N-PSDB; AAI61066.

XX Novel nucleic acids and polypeptides, useful for treating disorders such  
PT as central nervous system injuries.

XX Example 2; SEQ ID NO 6841; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form

|          |   |        |            |         |            |             |              |         |        |               |   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
|----------|---|--------|------------|---------|------------|-------------|--------------|---------|--------|---------------|---|-------|------------|-------|-----------|------|-------|------|---------|------|---------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|--------|---|---|---|---|---|---|---|---|---|---|---|--------|
| CC       | part of the printed specification   |        |            |         |            |             |              |         |        |               |   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| XX       | Sequence 1066 AA;   |        |            |         |            |             |              |         |        |               |   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| SQ       | Query Match 99.9%; Score 3468; DB 4; Length 1066;<br>Best Local Similarity 99.8%; Pred. No. 2.1e-269;<br>Matches 650; Conservative 0; Mismatches 1; Indels 0; Gaps 0;   |        |            |         |            |             |              |         |        |               |   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| QY       | 1   | MKNVPV | PVYCRPLVEK | DP      | TKMLWCAAGV | NLSGWRPNEDD | AGNGVKPAPGRD | PL      | TC     | DREGD 60      |   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| Db       | 416   | MKNVPV | PVYCRPLVEK | DP      | TKMLWCAAGV | NLSGWRPNEDD | AGNGVKPAPGRD | PL      | TC     | DREGD 475     |   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| QY       | 61  | GEPKSA | HTSPEKKK   | AKEL    | PENDATSSRV | WILTS       | TLTTSKV      | VIIDAN  | QPGTV  | VDQFTVCNA 120 |   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| Db       | 476   | GEPKSA | HTSPEKKK   | AKEL    | PENDATSSRV | WILTS       | TLTTSKV      | VIIDAN  | QPGTV  | VDQFTVCNA 535 |   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| QY       | 121   | HVLCIS | SIPAA      | SDSDYPP | GGEMFL     | DSVNPED     | PGADGVL      | AGITLV  | GCATRC | NVPRSNCS      | SR 180  |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| Db       | 536   | HVLCIS | SIPAA      | SDSDYPP | GGEMFL     | DSVNPED     | PGADGVL      | AGITLV  | GCATRC | NVPRSNCS      | SR 595  |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| QY       | 181   | GDTPLD | KGQGE      | VATIAN  | GKVNPSQ    | STEEATE     | VPDPGP       | SEPETAT | LRPG   | PLTEHVFT 240  |   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| Db       | 596   | GDTPLD | KGQGE      | VATIAN  | GKVNPSQ    | STEEATE     | VPDPGP       | SEPETAT | LRPG   | PLTEHVFT 655  |   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| QY       | 241   | DPAPT  | SSGPQ      | PGSENG  | PEPDS      | STRPEPE     | PSGDP        | TCAGSSA | APT    | TMWLG         | AQNGWLYVHSA 300   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| Db       | 656   | DPAPT  | SSGPQ      | PGSENG  | PEPDS      | STRPEPE     | PSGDP        | TCAGSSA | APT    | TMWLG         | AQNGWLYVHSA 715   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| QY       | 301   | VANWKK | CLHSI      | KLKDS   | SVLSLV     | HVKG        | RVLVAL       | ADGT    | LAI    | FHRG          | EDGQWDL   | SNYH  | MDLGHP 360 |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| Db       | 716   | VANWKK | CLHSI      | KLKDS   | SVLSLV     | HVKG        | RVLVAL       | ADGT    | LAI    | FHRG          | EDGQWDL   | SNYH  | MDLGHP 775 |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| QY       | 361   | HHSIR  | CMVV       | DRV     | WVW        | CGYK        | NKVH         | VIQ     | PKTMQ  | IEK           | SFDA  | HPRES | QVRL       | AWIGD | VWVSI 420 |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| Db       | 776   | HHSIR  | CMVV       | DRV     | WVW        | CGYK        | NKVH         | VIQ     | PKTMQ  | IEK           | SFDA  | HPRES | QVRL       | AWIGD | VWVSI 835 |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| QY       | 421   | RLD    | STL        | RLY     | HA         | HTH         | QHL          | QDV     | DI     | EPYV          | SKML  | GT    | KG         | LFS   | FVRIT     | ALLV | AGSRL | WVGT | NGV 480 |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| Db       | 836   | RLD    | STL        | RLY     | HA         | HTH         | QHL          | QDV     | DI     | EPYV          | SKML  | GT    | KG         | LFS   | FVRIT     | ALLV | AGSRL | WVGT | NGV 895 |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| QY       | 481   | VISIP  | L          | TET     | VL         | HR          | G            | OLL     | GL     | RANK          | TS  | PTSG  | EGAR       | PGGII | HVY       | GDD  | SSD   | RA   | ASS     | FIPY | CSM 540 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| Db       | 896   | VISIP  | L          | TET     | VL         | HR          | G            | OLL     | GL     | RANK          | TS  | PTSG  | EGAR       | PGGII | HVY       | GDD  | SSD   | RA   | ASS     | FIPY | CSM 955 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| QY       | 541   | AAQ    | LCF        | H       | CH         | R           | D            | AV      | K      | FFV           | S   | V     | P          | G     | N         | V    | L     | A    | T       | L    | N       | G | S | V | L | D | S | P | A | E | G | P | A | A | P | A | S      | E | V | E | G | Q | L | R | N | V | L | V | L 600  |
| Db       | 956   | AAQ    | LCF        | H       | CH         | R           | D            | AV      | K      | FFV           | S   | V     | P          | G     | N         | V    | L     | A    | T       | L    | N       | G | S | V | L | D | S | P | A | E | G | P | A | A | P | A | S      | E | V | E | G | Q | L | R | N | V | L | V | L 1015 |
| QY       | 601   | SGGE   | YID        | FR      | IG         | D           | ED           | DE      | TE     | EG            | AG  | D     | M          | S     | Q         | V    | K     | P    | V       | L    | S       | K | A | E | R | S | H | I | I | V | W | Q | S | Y | T | P | E 651  |   |   |   |   |   |   |   |   |   |   |   |        |
| Db       | 1016  | SGGE   | YID        | FR      | IG         | D           | ED           | DE      | TE     | EG            | AG  | D     | M          | S     | Q         | V    | K     | P    | V       | L    | S       | K | A | E | R | S | H | I | I | V | W | Q | S | Y | T | P | E 1066 |   |   |   |   |   |   |   |   |   |   |   |        |
| RESULT 3 |   |        |            |         |            |             |              |         |        |               | Human ORFX ORF2481 polypeptide sequence SEQ ID NO:4962.   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| AAB42717 |   |        |            |         |            |             |              |         |        |               | Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianaemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| ID       | AAB42717 standard; protein; 1069 AA.  |        |            |         |            |             |              |         |        |               | Query Match 99.9%; Score 3468; DB 3; Length 1069;<br>Best Local Similarity 99.8%; Pred. No. 2.1e-269;<br>Matches 650; Conservative 0; Mismatches 1; Indels 0; Gaps 0;   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| XX       | AAB42717;   |        |            |         |            |             |              |         |        |               | QY 1 MKNVPVYCRPLVEKDP   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| AC       | AAB42717;   |        |            |         |            |             |              |         |        |               | Db 419 MKNVPVYCRPLVEKDP   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| DT       | 08-FEB-2001 (first entry)   |        |            |         |            |             |              |         |        |               | QY 61 GEPKSAHTSPEKKKAKELPEMDATSSRVWILTS   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| XX       | Human ORFX ORF2481 polypeptide sequence SEQ ID NO:4962.   |        |            |         |            |             |              |         |        |               | Db 479 GEPKSAHASPEKKKAKELPEMDATSSRVWILTS  |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| XX       | Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianaemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; |        |            |         |            |             |              |         |        |               | QY 121 HVLCISSIPAAASDSYPPGGEMFLDSVNPEDPGADGVL   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| XX       | Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianaemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; |        |            |         |            |             |              |         |        |               | Db 539 HVLCISSIPAAASDSYPPGGEMFLDSVNPEDPGADGVL   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| XX       | Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianaemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; |        |            |         |            |             |              |         |        |               | QY 181 GDTPLD   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |
| XX       | Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianaemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; |        |            |         |            |             |              |         |        |               | Db 599 GDTPLD   |       |            |       |           |      |       |      |         |      |         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |        |   |   |   |   |   |   |   |   |   |   |   |        |

severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US008621.

31-MAR-1999; 99US-0127607P.

02-APR-1999; 99US-0127636P.

05-APR-1999; 99US-0127728P.

30-MAR-2000; 2000US-00540763.

(CURA-) CURAGEN CORP.

Shimkets RA, Leach M;

WPI; 2000-602362/57.

N-PSDB; AAC76926.

Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease.

Claim 11; Page 4131-4134; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic; anticonvulsant; antiarthritic; immunosuppressant; immunostimulant; cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive

QY 241 DPAPTPSSGPPQSGSENGPEPPSSSTRPEPEPSGDPGTGAGSSAAPTMMWLG AQNGWLYVHSA 300  
Db 659 DPAPTPSSGPPQSGSENGPEPPSSSTRPEPEPSGDPGTGAGSSAAPTMMWLG AQNGWLYVHSA 718  
QY 301 VANWKKCLHSIKLSDSVLSLVHVKGRLVALADGTLAIFHRGEDGQWDL SNYHLMDLGHP 360  
Db 719 VANWKKCLHSIKLSDSVLSLVHVKGRLVALADGTLAIFHRGEDGQWDL SNYHLMDLGHP 778  
QY 361 HHSIRCMVVYDRVWCGYKKNKHVIQPKTMQIEKSFDAHPRESQVRQLAWIGDVWVSI 420  
Db 779 HHSIRCMVVYDRVWCGYKKNKHVIQPKTMQIEKSFDAHPRESQVRQLAWIGDVWVSI 838  
QY 421 RLDSTLRLYHAHTHQHLQDVIDIEPVVSKMLGTGKLGFSFVRITALLVAGSRLWVGTNGV 480  
Db 839 RLDSTLRLYHAHTHQHLQDVIDIEPVVSKMLGTGKLGFSFVRITALLVAGSRLWVGTNGV 898  
QY 481 VISIPLTETVVLHKGQLLGLRANKTSPTSGEGARPGGIHVGDDSSDRAASSFIPYCSM 540  
Db 899 VISIPLTETVVLHKGQLLGLRANKTSPTSGEGARPGGIHVGDDSSDRAASSFIPYCSM 958  
QY 541 AQAQLCFHGHHRDAVKFFVSPGNVLATLNGSVLDSPAEGPGPAAPASEVEGQKLRNLVL 600  
Db 959 AQAQLCFHGHHRDAVKFFVSPGNVLATLNGSVLDSPAEGPGPAAPASEVEGQKLRNLVL 1018  
QY 601 SGGEGYIDFRIGDGEDDETEEGAGDMSQVKPVL SKAERSHIIVWQVSYTPE 651  
Db 1019 SGGEGYIDFRIGDGEDDETEEGAGDMSQVKPVL SKAERSHIIVWQVSYTPE 1069

RESULT 4

AAM40124  
ID AAM40124 standard; protein; 1014 AA.  
AC AAM40124;  
XX  
DT 22-OCT-2001 (first entry)  
DE Human polypeptide SEQ ID NO 3269.  
XX  
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
OS Homo sapiens.  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US034263.  
XX  
PR 23-DEC-1999; 99US-00471275.  
PR 21-JAN-2000; 2000US-00488725.  
PR 25-APR-2000; 2000US-00552317.  
PR 20-JUN-2000; 2000US-00598042.  
PR 19-JUL-2000; 2000US-00620312.  
PR 03-AUG-2000; 2000US-00653450.  
PR 14-SEP-2000; 2000US-00662191.  
PR 19-OCT-2000; 2000US-00693036.  
PR 29-NOV-2000; 2000US-00727344.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
PI Zhou P, Goodrich R, Drmanac RT;  
XX  
DR WPI; 2001-442253/47.  
DR N-PSDB; AAI59280.

XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders such  
PT as central nervous system injuries.  
XX  
PS Example 5; SEQ ID NO 3269; 10078pp; English.  
XX  
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification  
XX  
SQ Sequence 1014 AA;

Query Match 99.8%; Score 3465; DB 4; Length 1014;

Best Local Similarity 99.8%; Pred. No. 3.4e-269;

Matches 650; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKNVVPVYCRPLVEKDPMTKMLWCAAGVNLSGWRPNEDDAGNGVKPAPGRDPLTCDREGD 60  
Db 364 MKNVVPVYCRPLVEKDPMTKMLWCAAGVNLSGWRPNEDDAGNGVKPAPGRDPLTCDREGD 423  
QY 61 GEPKSAHTSPEKKKAKELPEMDATSSRVWILTSLTTSKVIIIDANQPGTVVDQFTVCNA 120  
Db 424 GEPKSAHTSPEKKKAKELPEMDATSSRVWILTSLTTSKVIIIDANQPGTVVDQFTVCNA 483  
QY 121 HVLCTISSIPAAASDSYPPGEMFLSDVNPEDPGADGVLAGITLVGCATRCNVPRNSCSR 180  
Db 484 HVLCTISSIPAAASDSYPPGEMFLSDVNPEDPGADGVLAGITLVGCATRCNVPRNSCSR 543  
QY 181 GDTPVLDKGQGEVATIANGKVNPSQSTEEATEATEVPDPGPPSEPETATLRPGPLTEHVFT 240  
Db 544 GDTPVLDKGQGEVATIANGKVNPSQSTEEATEATEVPDPGPPSEPETATLRPGPLTEHVFT 603  
QY 241 DPAPTPSSGPPQSGSENGPEPPSSSTRPEPEPSGDPGTGAGSSAAPTMMWLG AQNGWLYVHSA 300  
Db 604 DPAPTPSSGPPQSGSENGPEPPSSSTRPEPEPSGDPGTGAGSSAAPTMMWLG AQNGWLYVHSA 663  
QY 301 VANWKKCLHSIKLSDSVLSLVHVKGRLVALADGTLAIFHRGEDGQWDL SNYHLMDLGHP 360  
Db 664 VANWKKCLHSIKLSDSVLSLVHVKGRLVALADGTLAIFHRGEDGQWDL SNYHLMDLGHP 723  
QY 361 HHSIRCMVVYDRVWCGYKKNKHVIQPKTMQIEKSFDAHPRESQVRQLAWIGDVWVSI 420  
Db 724 HHSIRCMVVYDRVWCGYKKNKHVIQPKTMQIEKSFDAHPRESQVRQLAWIGDVWVSI 783  
QY 421 RLDSTLRLYHAHTHQHLQDVIDIEPVVSKMLGTGKLGFSFVRITALLVAGSRLWVGTNGV 480  
Db 784 RLDSTLRLYHAHTHQHLQDVIDIEPVVSKMLGTGKLGFSFVRITALLVAGSRLWVGTNGV 843  
QY 481 VISIPLTETVVLHKGQLLGLRANKTSPTSGEGARPGGIHVGDDSSDRAASSFIPYCSM 540  
Db 844 VISIPLTETVVLHKGQLLGLRANKTSPTSGEGARPGGIHVGDDSSDRAASSFIPYCSM 903  
QY 541 AQAQLCFHGHHRDAVKFFVSPGNVLATLNGSVLDSPAEGPGPAAPASEVEGQKLRNLVL 600  
Db 904 AQAQLCFHGHHRDAVKFFVSPGNVLATLNGSVLDSPAEGPGPAAPASEVEGQKLRNLVL 963  
QY 601 SGGEGYIDFRIGDGEDDETEEGAGDMSQVKPVL SKAERSHIIVWQVSYTPE 651  
Db 964 SGGEGYIDFRIGDGEDDETEEGAGDMSQVKPVL SKAERSHIIVWQVSYTPE 1014

|  |
|--|
| RESULT 5   |
| ABG09904   |
| ID ABG09904 standard; protein; 1069 AA.                                      |
| XX AC ABG09904;  |
| XX DT 13-FEB-2002 (first entry)  |
| XX DE Novel human diagnostic protein #9895.                                  |
| XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;       |
| XX KW food supplement; medical imaging; diagnostic; genetic disorder.        |
| XX OS Homo sapiens.  |
| XX PN WO200175067-A2.  |
| XX PD 11-OCT-2001.   |
| XX PF 30-MAR-2001; 2001WO-US008631.  |
| XX PR 31-MAR-2000; 2000US-00540217.  |
| XX PR 23-AUG-2000; 2000US-00649167.  |
| XX PA (HYSE-) HYSEQ INC.   |
| XX PI Drmanac RT, Liu C, Tang YT;  |
| XX DR WPI; 2001-639362/73.   |
| DR N-PSDB; AAS74091.   |
| XX New isolated polynucleotide and encoded polypeptides, useful in           |
| PT diagnostics, forensics, gene mapping, identification of mutations         |
| PT responsible for genetic disorders or other traits and to assess           |
| PT biodiversity.   |
| XX Claim 20; SEQ ID NO 40263; 103pp; English.                                |
| PS The invention relates to isolated polynucleotide (I) and polypeptide (II) |
| XX sequences. (I) is useful as hybridisation probes, polymerase chain        |
| CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,   |
| CC and in recombinant production of (II). The polynucleotides are also used  |
| CC in diagnostics as expressed sequence tags for identifying expressed       |
| CC genes. (I) is useful in gene therapy techniques to restore normal         |
| CC activity of (II) or to treat disease states involving (II). (II) is       |
| CC useful for generating antibodies against it, detecting or quantitating a  |
| CC polypeptide in tissue, as molecular weight markers and as a food          |
| CC supplement. (II) and its binding partners are useful in medical imaging   |
| CC of sites expressing (II). (I) and (II) are useful for treating disorders  |
| CC involving aberrant protein expression or biological activities. The       |
| CC polypeptide and polynucleotide sequences have applications in             |
| CC diagnostics, forensics, gene mapping, identification of mutations         |
| CC responsible for genetic disorders or other traits to assess biodiversity  |
| CC and to produce other types of data and products dependent on DNA and      |
| CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic   |
| CC amino acid sequences of the invention. Note: The sequence data for this   |
| CC patent did not appear in the printed specification, but was obtained in   |
| CC electronic format directly from WIPO at                                   |
| XX ftp.wipo.int/pub/published_pct_sequences                                  |
| SQ Sequence 1069 AA;   |
| Query Match 97.0%; Score 3369; DB 4; Length 1069;                            |
| Best Local Similarity 97.7%; Pred. No. 1.9e-261;                             |
| Matches 636; Conservative 1; Mismatches 14; Indels 0; Gaps 0                 |
| Qy 1 MKNVPVPPYCRPLVEKDPTMKLWCAAGVNLSGRPNEDDAGNGVKPAPGRDPLTCDREGD 60          |
| Dd                   :   |
| 419 MKNVPVPPYCRPSVGKDPITLCVAPAANLSGRPNEDDAGNGVKPAPGHGLTSDREGD 478            |
| Qy 61 GEPKSAHTSPKKAKELPENDATSSRWILSTLTTSKVVIIDANQPVTVDQFTVCNA 120            |
| Dd   |
| 479 GEPKSAHTSPKKAKELPENDATSSRWILSTLTTSKVVIIDANQPVTVDQFTVCNA 538              |



QY 241 DPAPTPSSGPPQSGSENGPEPDSSSTRPEPEPSGDPGTGAGSSAAPTMLGAQNGWLYVHSA 300  
Db 904 DPAPTPSSSTQPASENGSENGTIVQPQVPSGELSTTTSSAAPTMLGAQNGWLYVHSA 963  
QY 301 VANWKKCLHSIKLSDSVLSLVHVKGRVLVALADGTLAIFHRGEDGQWDLNSYHLMDLGHP 360  
Db 964 VANWKKCLHSIKLSDSVLSLVHVKGRVLVALADGTLAIFHRGEDGQWDLNSYHLMDLGHP 1023  
QY 361 HHSIRCMVAVVYDRVWCGYKKNKVHVIOQPKTMQIEKSFDAHPRRESQVRLAWIGDVWVSI 420  
Db 1024 HHSIRCMVAVVNDRVWCGYKKNKVHVIOQPKTMQIEKSFDAHPRRESQVRLAWIGDVWVSI 1083  
QY 421 RLDSTLRLYHAHTHQHLQDVIEPVVSKMLGTGKLGFSFVRITALLVAGSRLWVGTVNGV 480  
Db 1084 RLDSTLRLYHAHTHQHLQDVIEPVVSKMLGTGKLGFSFVRITALLIAGNRLWVGTVNGV 1143  
QY 481 VISIPLTETVVLHRRGQLLGLRANKTSPTSGEGARPGGIIHVYGGDSSDRAASSFIPYCSM 540  
Db 1144 VISIPLTETVVLHRRGQLLGLRANKTSPTSGEGTRPGGIIHVYGGDSSDKAASSFIPYCSM 1203  
QY 541 AQAQLCFHGHHRDAVKFFVSVPGNVLATLNGSVLDSPAEGPGPAAPASEVEGQKLRNLVL 600  
Db 1204 AQAQLCFHGHHRDAVKFFVSVPGNVLATLNGSVLDSPSEGGPGPAAPADAEGQKLNALVL 1263  
QY 601 SGGEGYIDFRIGDGEDDETEEGAGDMSQVKPVLSKAERSHIIVWQVSYTPE 651  
Db 1264 SGGEGYIDFRIGDGEDDETEECAGDVNQTKPSLSKAERSHIIVWQVSYTPE 1314

RESULT 8  
AAB12878  
ID AAB12878 standard; protein; 1336 AA.  
XX  
AC AAB12878;  
XX  
DT 01-NOV-2000 (first entry)  
XX  
DE Murine JNK3 binding protein amino acid sequence #4.  
XX  
KW JNK3; Jun N-terminal protein kinase 3; JNK3 binding protein; JSAP;  
KW JNK/SAPK associated protein; neurotropic; neuroprotective; antiinflammatory;  
KW antiparkinsonian; anticonvulsant; cerebroprotective; antiinflammatory;  
KW nervous system disease; Alzheimer's disease; Parkinson's disease;  
KW Huntington's disease; multiple sclerosis; muscular dystrophy; apoplexy;  
KW schizophrenia; drowsy; inflammatory disorders; mouse.

Mus sp.  
XX  
XX WO200031132-A1.  
XX  
XX 02-JUN-2000.  
XX  
PF 19-NOV-1999; 99WO-JP006487.  
XX  
PR 24-NOV-1998; 98JP-00332484.  
PR 02-SEP-1999; 99JP-00248442.  
XX  
PA (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
XX Ichimura M, Hirose R, Yoshioka K;  
PI  
XX WPI; 2000-400042/34.  
DR N-PSDB; AAA62902.

Polypeptides binding to Jun N-terminal protein kinase for treatment and diagnosis of nervous system and inflammatory disorders.

Claim 1; Page 220-230; 254pp; Japanese.  
XX  
CC The invention relates to polypeptides which bind to JNK3 (Jun N-terminal protein kinase 3). The present sequence represents a JNK3 binding protein (JSAP; JNK/SAPK associated protein) amino acid sequence. Included in the invention are polypeptides derived from the JNK3 binding proteins, DNA

CC sequences encoding the proteins, expression vectors containing the DNA, and host cells transformed by the vectors. The polypeptides exhibit neurotropic, neuroprotective, antiparkinsonian, anticonvulsant, cerebroprotective, neuropleptic, and antiinflammatory activity. JNK3 binding proteins are used in the treatment, prevention and diagnosis of diseases associated with the JNK3 cascade, such as diseases of the nervous system including Alzheimer's disease, Parkinson's disease, Huntington's disease, multiple sclerosis, muscular dystrophy, apoplexy and schizophrenia, drowsy, and inflammatory disorders

XX SQ Sequence 1336 AA;

Query Match 91.9%; Score 3191; DB 3; Length 1336;  
Best Local Similarity 91.2%; Pred. No. 5.3e-247;  
Matches 594; Conservative 24; Mismatches 33; Indels 0; Gaps 0;

QY 1 MKNVPVPVYCRPLVEKDPMTKMLWCAAGVNLGWRPNEDDAGNGVKPAPGRDPLTCDREGD 60  
Db 686 MKNVPVPVYCRPLVEKDPSTKLWCAAGVNLGWRPNEDDAGNGVKPAPGRDPLTCDREGD 745  
QY 61 GEPKSAHTSPEKKKAKELPEMDATSSRVWILTSTLTTSKVVIIDANQPGTIVDQFTVCNA 120  
Db 746 GEPKSTHPSPEKKKAKETPEADATSSRVWILTSTLTTSKVVIIDANQPGTIVDQFTVCNA 805  
QY 121 HVLCISSIPAASDSYPPGEMFLDSVNPEPDGADGVLGITLVGCATRCNVRNSCSSR 180  
Db 806 HVLCISSIPAASDSYPPGEMFLDSVNPEPDGADGVLGITLVGCATRCNVRNSCSSR 865  
QY 181 GDTPLVDKGGEVATIANGVNPSQSTEEATEATEVDPGSEPEPETATLRPGPLTEHVFT 240  
Db 866 GDTPLVDKGGDVATTANGVNPVSQSTEEATEATEVDPGSESEATTVRPGPLTEHVFT 925  
QY 241 DPAPTPSSGPPQSGSENGPEPDSSSTRPEPEPSGDPGTGAGSSAAPTMLGAQNGWLYVHSA 300  
Db 926 DPAPTPSSSTQPASENGSENGTIVQPQVPSGELSTTTSSAAPTMLGAQNGWLYVHSA 985  
QY 301 VANWKKCLHSIKLSDSVLSLVHVKGRVLVALADGTLAIFHRGEDGQWDLNSYHLMDLGHP 360  
Db 986 VANWKKCLHSIKLSDSVLSLVHVKGRVLVALADGTLAIFHRGEDGQWDLNSYHLMDLGHP 1045  
QY 361 HHSIRCMVAVVYDRVWCGYKKNKVHVIOQPKTMQIEKSFDAHPRRESQVRLAWIGDVWVSI 420  
Db 1046 HHSIRCMVAVVNDRVWCGYKKNKVHVIOQPKTMQIEKSFDAHPRRESQVRLAWIGDVWVSI 1105  
QY 421 RLDSTLRLYHAHTHQHLQDVIEPVVSKMLGTGKLGFSFVRITALLVAGSRLWVGTVNGV 480  
Db 1106 RLDSTLRLYHAHTHQHLQDVIEPVVSKMLGTGKLGFSFVRITALLIAGNRLWVGTVNGV 1165  
QY 481 VISIPLTETVVLHRRGQLLGLRANKTSPTSGEGARPGGIIHVYGGDSSDRAASSFIPYCSM 540  
Db 1166 VISIPLTETVVLHRRGQLLGLRANKTSPTSGEGTRPGGIIHVYGGDSSDKAASSFIPYCSM 1225  
QY 541 AQAQLCFHGHHRDAVKFFVSVPGNVLATLNGSVLDSPAEGPGPAAPASEVEGQKLRNLVL 600  
Db 1226 AQAQLCFHGHHRDAVKFFVSVPGNVLATLNGSVLDSPSEGGPGPAAPADAEGQKLNALVL 1285  
QY 601 SGGEGYIDFRIGDGEDDETEEGAGDMSQVKPVLSKAERSHIIVWQVSYTPE 651  
Db 1286 SGGEGYIDFRIGDGEDDETEECAGDVNQTKPSLSKAERSHIIVWQVSYTPE 1336

RESULT 9  
AAB12877  
ID AAB12877 standard; protein; 1337 AA.  
XX  
AC AAB12877;  
XX  
DT 01-NOV-2000 (first entry)

XX Murine JNK3 binding protein amino acid sequence #3.  
CC JNK3; Jun N-terminal protein kinase 3; JNK3 binding protein; JSAP;  
CC JNK/SAPK associated protein; neurotropic; neuroprotective; neuropleptic;

KW antiparkinsonian; anticonvulsant; cerebroprotective; antiinflammatory;  
KW nervous system disease; Alzheimer's disease; Parkinson's disease;  
KW Huntington's disease; multiple sclerosis; muscular dystrophy; apoplexy;  
KW schizophrenia; dropsy; inflammatory disorders; mouse.  
XX  
OS Mus sp.  
XX  
XX WO200031132-A1.  
PN  
XX  
XX  
XX PD 02-JUN-2000.  
XX  
XX  
XX PF 19-NOV-1999; 99WO-JP006487.  
XX  
XX PR 24-NOV-1998; 98JP-00332484.  
PR 02-SEP-1999; 99JP-00248442.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
PA  
XX Ichimura M, Hirose R, Yoshioka K;  
PI  
XX WPI; 2000-400042/34.  
DR N-PSDB; AAA62901.  
DR  
XX  
XX Polypeptides binding to Jun N-terminal protein kinase for treatment and  
PT diagnosis of nervous system and inflammatory disorders.  
PT  
XX  
XX  
PS Claim 1; Page 210-220; 254pp; Japanese.  
XX  
XX The invention relates to polypeptides which bind to JNK3 (Jun N-terminal  
CC protein kinase 3). The present sequence represents a JNK3 binding protein  
CC (JSAP; JNK/SAPK associated protein) amino acid sequence. Included in the  
CC invention are polypeptides derived from the JNK3 binding proteins, DNA  
CC sequences encoding the proteins, expression vectors containing the DNA,  
CC and host cells transformed by the vectors. The polypeptides exhibit  
CC neurotropic, neuroprotective, antiparkinsonian, anticonvulsant,  
CC cerebroprotective, neuroleptic, and antiinflammatory activity. JNK3  
CC binding proteins are used in the treatment, prevention and diagnosis of  
CC diseases associated with the JNK3 cascade, such as diseases of the  
CC nervous system including Alzheimer's disease, Parkinson's disease,  
CC Huntington's disease, multiple sclerosis, muscular dystrophy, apoplexy  
CC and schizophrenia, dropsy, and inflammatory disorders  
XX  
SQ Sequence 1337 AA;  
  
Query Match 91.9%; Score 3191; DB 3; Length 1337;  
Best Local Similarity 91.2%; Pred. No. 5.3e-247;  
Matches 594; Conservative 24; Mismatches 33; Indels 0; Gaps 0;  
  
QY 1 MKNVPVPVYCRPLVEKDPSTKLCWCAAGVNLGWRPNEDDAGNVKPAAPGRDPLTCDREGD 60  
Db 687 MKNVPVPVYCRPLVEKDPSTKLCWCAAGVNLGWRPNEDDAGNVKPAAPGRDPLTCDREGD 746  
  
QY 61 GEPKSAHTSPEKKAKELPEMDATSSRVWILTSTLTTSKWVLIIDANOPGTVDQFTVCNA 120  
Db 747 GEPKSTHPSPEKKAKETPEADATSSRVWILTSTLTTSKWVLIIDANOPGTIVDQFTVCNA 806  
  
QY 121 HVLCTSSIPAASDSYPPGEMFLDSVNPEDPGADGVLAGITLVGCATRCNVPRSNCSRR 180  
Db 807 HVLCTSSIPAASDSYPPGEMFLDSVNPEDSGADGVLAGITLVGCATRCNVPRSNCSRR 866  
  
QY 181 GDTPVLDKGQGEVATIANGKVNPSQSTEEATEATEVPDPGPSEPETATLRPGPLTEHVFT 240  
Db 867 GDTPVLDKGQGDVATTANGKVNPSQSTEEATEATEVPDPGPSESEATTVRPGPLTEHVFT 926  
  
QY 241 DPAPTPSSGPQPGSENGPEBPDSSSTRPEPPSGDPTGAGSSAAPTMLGAQNGWLTVHSA 300  
Db 927 DPAPTPSSSTQPASENGSENGTIVQPQVPSGELSTTSSAAPTMLGAQNGWLTVHSA 986  
  
QY 301 VANWKKCLHSIKLKDSVLSLVHVHVKGRVLVALADGTTAIFHRGEDGQWDLSNYHLMDLGHP 360  
Db 987 VANWKKCLHSIKLKDSVLSLVHVHVKGRVLVALADGTTAIFHRGEDGQWDLSNYHLMDLGHP 1046  
  
QY 361 HHSIRCMVYDRVWCGYKKNKVHVVIQPKTMQIEKSFDAHPRRESQVRQLAWIGDGVWVSI 420

Db 1047 HHSIRCMVYDRVWCGYKKNKVHVVIQPKTMQIEKSFDAHPRRESQVRQLAWIGDGVWVSI 1106  
QY 421 RLDSTLRLYHAHTHQHLQDVLDIEPYVSKMLGTGKLGFSFVRITALLVAGSRLWVGTVNGV 480  
Db 1107 RLDSTLRLYHAHTHQHLQDVLDIEPYVSKMLGTGKLGFSFVRITALLIAGNRLWVGTVNGV 1166  
QY 481 VISIPLTETVVLHRGQLLGLRANKTSPTSGEGARPGGIHVYGGDDSSDRAASSFIPYCSM 540  
Db 1167 VISIPLTETVVLHRGQLLGLRANKTSPTSGEGTRPGGIHVYGGDDSSDKAASSFIPYCSM 1226  
QY 541 AQAQLCFHGHHRDAVKFFVSPGVNVLATLNGSVLDSPAEAGPGPAAPASEVEGQKLRNLVL 600  
Db 1227 AQAQLCFHGHHRDAVKFFVSPGVNVLATLNGSVLDSPSEGGPGPAAPADAEGQKLRNALVL 1286  
QY 601 SGGEGYIDFRIGDGEDDETEEGAGDMSQVKPVLKAERSHIIVMQVSYTPE 651  
Db 1287 SGGEGYIDFRIGDGEDDETEECAGDVNQTKPSLSKAERSHIIVMQVSYTPE 1337  
  
RESULT 10  
ABG09902  
ID ABG09902 standard; protein; 809 AA.  
XX  
AC ABG09902;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #9893.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US008631.  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX (HYSE-) HYSEQ INC.  
PA Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
DR N-PSDB; AAS74089.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 20; SEQ ID NO 40261; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 809 AA;

Query Match 79.4%; Score 2757.5; DB 4; Length 809;  
Best Local Similarity 93.2%; Pred. No. 1.8e-212;  
Matches 524; Conservative 1; Mismatches 2; Indels 35; Gaps 1;

QY 125 ISSIPAAASDSDYPPGEMFLDSVNPEDPGADGVLAGITLVGCATRCNVPRSNCSRGDTP 184  
Db :|||||  
QY 248 LSLSPAASDSDYPPGEMFLDSVNPEDPGADGVLAGITLVGCATRCNVPRSNCSRGDTP 307  
Db LSLSPAASDSDYPPGEMFLDSVNPEDPGADGVLAGITLVGCATRCNVPRSNCSRGDTP 307

QY 185 VLDKGQGEVATIANGKVNPSQSTEEATEATEVDPGPSEPETATLRPGPLTEHVFTDPAP 244  
Db |||||  
QY 308 VLDKGQGEVATIANGKVNPSQSTEEATEATEVDPGPSEPETATLRPGPLTEHVFTDPAP 367  
Db VLDKGQGEVATIANGKVNPSQSTEEATEATEVDPGPSEPETATLRPGPLTEHVFTDPAP 367

QY 245 TPSSGPQPGSENGEPDSSSTRPEPEPSGDPPTGAGSSAAPTMMLGQNGWLYVHSAVANW 304  
Db |||||  
QY 368 TPSSGPQPGSENGEPDSSSTRPEPEPSGDPPTGAGSSAAPTMMLGQNGWLYVHSAVANW 427  
Db TPSSGPQPGSENGEPDSSSTRPEPEPSGDPPTGAGSSAAPTMMLGQNGWLYVHSAVANW 427

QY 305 KKCLHSIKLSDSLVHVKGRLVALADGTTLAIFHRGEDGOWDLSNYHLMDLGHPPHSI 364  
Db |||||  
QY 428 KKCLHSIKLSDSLVHVKGRLVALADGTTLAIFHRGEDGOWDLSNYHLMDLGHPPHSI 487  
Db KKCLHSIKLSDSLVHVKGRLVALADGTTLAIFHRGEDGOWDLSNYHLMDLGHPPHSI 487

QY 365 RCMAVVYDRVWCYKKNKHVIOPKTMQIE----- 393  
Db |||||

QY 488 RCMAVVYDRVWCYKKNKHVIOPKTMQIEASAGQPGEGKRLLLASSSPASSIRKSTRPT 547  
Db RCMAVVYDRVWCYKKNKHVIOPKTMQIEASAGQPGEGKRLLLASSSPASSIRKSTRPT 547

QY 394 -----KSFDAHPRESQVRQLAWIGDGVVWSIRLDSITRLRYHAHTHQHLQDVDPYVSKM 449  
Db |||||

QY 548 HDSTKSFDAHPRESQVRQLAWIGDGVVWSIRLDSITRLRYHAHTHQHLQDVDPYVSKM 607  
Db HDSTKSFDAHPRESQVRQLAWIGDGVVWSIRLDSITRLRYHAHTHQHLQDVDPYVSKM 607

QY 450 LGTGKLGFSFVRITALLVAGSRLWVGTVGNGVVISIPLTETVVLHKGQLGLRANKTSPTS 509  
Db |||||

QY 608 LGTGKLGFSFVRITALLVAGSRLWVGTVGNGVVISIPLTETVVLHKGQLGLRANKTSPTS 667  
Db LGTGKLGFSFVRITALLVAGSRLWVGTVGNGVVISIPLTETVVLHKGQLGLRANKTSPTS 667

QY 510 GEGARPGGIIHVYGGDSSDRAASSFIPYCSMAQAQLCFHGHDAVKFFVSPGNVLATLN 569  
Db |||||

QY 668 GEGARPGGIIHVYGGDSSDRAASSFIPYCSMAQAQLCFHGHDAVKFFVSPGNVLATLN 727  
Db GEGARPGGIIHVYGGDSSDRAASSFIPYCSMAQAQLCFHGHDAVKFFVSPGNVLATLN 727

QY 570 GSVLDSPAEGPAPAPASEVEGQKLRNVLSGGEGYIDFRIGDGEDDETEEGAGDMSQV 629  
Db GSVLDSPAEGPAPAPASEVEGQKLRNVLSGGEGYIDFRIGDGEDDETEEGAGDMSQV 629

QY 728 GSVLDSPAEGPAPAPASEVEGQKLRNVLSGGEGYIDFRIGDGEDDETEEGAGDMSQV 787  
Db GSVLDSPAEGPAPAPASEVEGQKLRNVLSGGEGYIDFRIGDGEDDETEEGAGDMSQV 787

QY 630 KPVLSKAERSHIIVWQVSYTPE 651  
Db KPVLSKAERSHIIVWQVSYTPE 809

RESULT 11  
ADA45192

ID ADA45192 standard; protein; 1307 AA.  
XX  
AC ADA45192;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Murine JLP.  
XX

KW Cytostatic; Gene therapy; murine; scaffolding protein; JLP;  
KW JNK-associated Leucine zipper Protein; MEK kinase 3; MEK3;  
KW MAP kinase 4; MKK4; c-Jun NH2-terminal kinase; JNK;  
KW p38 MAP kinase; MAPK; c-Myc; MAX; apoptosis; cancer.  
XX  
OS Mus musculus.  
XX  
PN WO2003066652-A2.  
XX

PD 14-AUG-2003.  
XX  
PF 04-FEB-2003; 2003WO-US003355.  
XX  
PR 05-FEB-2002; 2002US-0354377P.  
XX  
PA (UTEM ) UNIV TEMPLE.  
XX  
PI Lee CM, Dhanasekaran N, Reddy PE;  
XX  
DR WPI; 2003-731487/69.  
XX  
DR N-PSDB; ADA45191.  
XX  
PT New scaffolding nucleic acid sequences, designated as JLP, useful for  
PT modulating apoptotic response in a cell, and thus for treating metastatic  
PT cancer.  
XX  
PS Claim 22; Page 80-83; 102pp; English.  
XX  
CC The present invention relates to novel human and murine scaffolding  
CC proteins, JLP (for JNK-associated Leucine zipper Protein, ADA45190 and  
CC ADA45192). JLP tethers MEK kinase 3 (MEK3), Mitogen-Activated Protein  
CC (MAP) kinase 4 (MKK4), c-Jun NH2-terminal kinase (JNK), p38 MAP  
CC kinase (MAPK), c-Myc and MAX into a signalling module which controls the  
CC apoptotic response. JLP therefore functions as a signalling conduit to  
CC transmit extracellular signals to the nucleus through MEK3-MKK4-  
CC JNK/p38/MAPK/c-Myc/MAX signalling module. The JLP sequences are useful  
CC for modulating apoptotic response in a cell, and thus for treating  
CC metastatic cancer.  
XX  
SQ Sequence 1307 AA;

Query Match 52.4%; Score 1821.5; DB 6; Length 1307;  
Best Local Similarity 56.4%; Pred. No. 6.1e-137;  
Matches 381; Conservative 93; Mismatches 147; Indels 55; Gaps 18;

QY 1 MKNVPVPVYCRPLVEKDPMTMKLWCAAGVNLGWRPNEDDAGNGVKPAP-GRDPLTCDREG 59  
Db |||||  
QY 652 MKNLPVPVYLRPLDEKDSMKLWCAVGNLSGGKTR--DGSVVVGASVFYKDIAGLDTEG 709  
Db MKNLPVPVYLRPLDEKDSMKLWCAVGNLSGGKTR--DGSVVVGASVFYKDIAGLDTEG 709

QY 60 DGEPKSAHTSPE-----KKKAKELPEMDATSSRVWILTSTLTTSKVVIDANQPGTVVD 113  
Db :|||  
QY 710 SKQRSASQSSLDKLDQELKEQKEFKNQELSSQVWICTSTHTTKVIIIDAVQPGNILD 769  
Db SKQRSASQSSLDKLDQELKEQKEFKNQELSSQVWICTSTHTTKVIIIDAVQPGNILD 769

QY 114 QFTVCNAHVLCISSIPAAASDSDYPPGEMFLDS-----VNPEDPGADGVLAGIT 162  
Db |||||  
QY 770 SFTVCNSHVLCIASVPGARETDYPAGEELSESGQVDKASLCGSMTSNSSAEMDSLGGIT 829  
Db SFTVCNSHVLCIASVPGARETDYPAGEELSESGQVDKASLCGSMTSNSSAEMDSLGGIT 829

QY 163 LVGCATR-CNVPRSNCSRGDTPVLDKGQGEVATIANGKVNPSQST-EEATEATEVDPGP 220  
Db :|||  
QY 830 VVGCTEGTLGAATSPSTNGASPVIEK-PPMET-ENSEVDENIPTAEATEATE-GNAG 886  
Db VVGCTEGTLGAATSPSTNGASPVIEK-PPMET-ENSEVDENIPTAEATEATE-GNAG 886

QY 221 PSEPETATLRPGPLTEHVFTDPAPT-----SSGPQSGSENGPEPDSSSTRP-EPEPSGD 274  
Db :|||  
QY 887 STEDTVDISQPGVYTEHVFTDPLGVQIPEDLSPVFQSSNDSVYKQISVLPNEQDLARE 946  
Db STEDTVDISQPGVYTEHVFTDPLGVQIPEDLSPVFQSSNDSVYKQISVLPNEQDLARE 946

QY 275 PTGAGSSAAPTMMLGQNGWLYVHSAVANWKKCLHSIKLSDSLVHVKGRLVALADG 334  
Db |||||

QY 947 EAQKMSSLLPTMWLGQNGCLYVHSSVAQWRKCLHSIKLSDSLVHVKGRLVALADG 1006  
Db EAQKMSSLLPTMWLGQNGCLYVHSSVAQWRKCLHSIKLSDSLVHVKGRLVALADG 1006

QY 335 TLAI FHRGEDGOWDLSNYHLMDLGHPPHSIRCMNAVVDVRCVCGYKKNKHVIOPKTMQIEK 394  
Db |||||

QY 1007 TLAI FHRGVGGQWDLNLYHLMDLGRPHHSIRCMVTVDKVCYGRNKHVYVQPKAMKIEK 1066  
Db TLAI FHRGVGGQWDLNLYHLMDLGRPHHSIRCMVTVDKVCYGRNKHVYVQPKAMKIEK 1066

QY 395 SFDAHPRESQVRQLAWIGDGVVWSIRLDSITRLRYHAHTHQHLQDVDPYVSKMLGTGK 454  
Db |||||

QY 1067 SFDAHPRESQVRQLAWIGDGVVWSIRLDSITRLRYHAHTHQHLQDVDPYVSKMLGTGK 1126  
Db SFDAHPRESQVRQLAWIGDGVVWSIRLDSITRLRYHAHTHQHLQDVDPYVSKMLGTGK 1126

QY 455 LGFSFVRITALLVAGSRLWVGTVGNGVVISIPLTETVVLHKGQLGLRANKTSPTS GEGAR 514  
Db |||||

QY 1127 LGFSFVRITALLVAGSRLWVGTVGNGVVISIPLTET-----NKTSGTGPN--R 1171  
Db LGFSFVRITALLVAGSRLWVGTVGNGVVISIPLTET-----NKTSGTGPN--R 1171

QY 515 PGGIIHVYGGDSSDRAA-SSFIPYCSMAQAQLCFHGHDAVKFFVSPGNVL---ATLNG 570  
Db PGGIIHVYGGDSSDRAA-SSFIPYCSMAQAQLCFHGHDAVKFFVSPGNVL---ATLNG 570





CC ADA45192). JLP tethers MEK kinase 3 (MEKK3), Mitogen-Activated Protein  
CC (MAP) kinase kinase 4 (MKK4), c-Jun NH2-terminal kinase (JNK), p38 MAP  
CC kinase (WAPK), c-Myc and MAX into a signalling module which controls the  
CC apoptotic response. JLP therefore functions as a signalling conduit to  
CC transmit extracellular signals to the nucleus through MEKK3-MKK4-  
CC JNK/p38/MAPK/c-Myc/MAX signalling module. The JLP sequences are useful  
CC for modulating apoptotic response in a cell, and thus for treating  
CC metastatic cancer. The present sequence is a C-terminal deletion mutant  
CC of murine JLP, generated in an example from the invention.

XX SQ Sequence 1165 AA;

Query Match 42.2%; Score 1467; DB 6; Length 1165;  
Best Local Similarity 57.3%; Pred. No. 1.6e-108;  
Matches 305; Conservative 68; Mismatches 115; Indels 44; Gaps 12;

QY 1 MKNVPVPVYCRPLVEKDPMTKMLWCAAGVNLGWRPNEDDAGNGVKPAP-GRDPLTCDREG 59  
Db 652 MKNLPVPVYLRPLDEKDSMLWCAVGNLSGGKTR--DGGSVVGASVFYKDIAGLDTEG 709

QY 60 DGEPKSAHTSPE-----KKAKELPEMDATSGRWILSTLTTSKVVIIDANQPGTVVD 113  
Db 710 SKQRSASQSSLDKLDQELKEQKFKNQEESSQVWICTSTHSTTKVVIIDAVQPGNILD 769

QY 114 QFTVCNAHVLCISSIPAAASDSYPPGEMFLDSD-----VNPEDPGADGVLAGIT 162  
Db 770 SFTVCNSHVLCIASVPGARETDYPAGEELSESGQVDKASLCGSMTSNSSAEMDSLGGIT 829

QY 163 LVGCATR-CNVPRNSCSSRGDTPVLDKGQGEVATANGKVNPSQST-EEATEATEVPDPG 220  
Db 830 VVGCSSTGLTGAATSPSTNGASPVIEK-PPEMET-ENSEVDENIPTAEATEATE-ENAG 886

QY 221 PSEPETATLRPGPLTEHVFTDPAPTP-----SSGPQPGSENGPEPDSSSTRP-EPEPSGD 274  
Db 887 STEDTVDISQPGVYTEHVFTDPLGVQIPEDLSPVFQSSNDSVYKDQISVLPNEQDLARE 946

QY 275 PTGAGSSAAPTMMWLGQNGWLYVHSAVANWKKCLHSIKLKDSVLSLVHVKGRLVALADG 334  
Db 947 EAQKMSLLPTMWLGQNGCLYVHSSVAQWRKCLHSIKLKDSILSVHVKGIVLVALADG 1006

QY 335 TLAIHFRGEDGQWDLNSYHLMOLGHPHHSIRCMVYDVRWCYKKNKVHVIOPKTMQIEK 394  
Db 1007 TLAIHFRGVDGQWDLNSYHLLDLGRPHHSIRCMTVVHDKWCGYRNKIYVVPQKAMKIEK 1066

QY 395 SFDAPRRRESQVRQLAWIGDGVVWSIRLDSSTLRLYHAHTHQHLQDVVDIEPYVSKMLGTGK 454  
Db 1067 SFDAPRKESQVRQLAWVGDVVWSIRLDSSTLRLYHAHTYQHLQDVVDIEPYVSKMLGTGK 1126

QY 455 LGFSFVRITALLVAGSRLWVGTVGNGVVISIPLTETVVLHRGQLGLRANKTS 506  
Db 1127 LGFSFVRITALLMVSCNRLWVGTVGNGVVISIPLTET-----NKTS 1165

RESULT 15  
AAW64473  
XX ID AAW64473 standard; protein; 828 AA.  
AC AAW64473;

XX DT 16-OCT-1998 (first entry)  
XX DE Human secreted protein from clone EC172\_1.  
XX KW Secreted protein; nutrition; cytokine; cell proliferation; activin;  
KW differentiation; immune system; stimulator; suppressor; tissue growth;  
KW haematopoiesis regulation; inhibin; chemotactic; chemokinetic; cadherin;  
KW haemostatic; thrombolysis; receptor; ligand; anti-inflammatory.

XX OS Homo sapiens.  
XX PN WO9830589-A2.  
XX PD 16-JUL-1998.

XX 20-DEC-1997; 97WO-US023506.  
XX 10-JAN-1997; 97US-00781225.  
XX 18-DEC-1997; 97US-00993228.  
XX (GEMY ) GENETICS INST INC.  
XX Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;  
PI Spaulding V, Agostino MJ;  
XX WPI: 1998-413682/35.  
DR N-PSDB; AAV46322.  
XX New isolated nucleic acids and secreted proteins - isolated from human  
PT foetal kidney, adult placenta, adult colon, adult testes, foetal brain  
PT and adult brain cDNA libraries.  
XX Claim 39a; Page 97-100; 122pp; English.  
XX This sequence represents a novel secreted protein from clone EC172\_1  
CC isolated from a human adult brain cDNA library. This novel protein may  
CC have biological activity, e.g. nutritional activity, cytokine and cell  
CC proliferation/differentiation activity, immune stimulating or suppressing  
CC activity, haematopoiesis regulating activity, tissue growth activity,  
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
XX activity or other activities  
SQ Sequence 828 AA;

Query Match 41.6%; Score 1446.5; DB 2; Length 828;  
Best Local Similarity 57.7%; Pred. No. 4.5e-107;  
Matches 300; Conservative 65; Mismatches 124; Indels 31; Gaps 10;  
QY 1 MKNVPVPVYCRPLVEKDPMTKMLWCAAGVNLGWRPNEDDAGNGVKPAP-GRDPLTCDREG 59  
Db 314 MKNLPVPVYLRPLDEKDTSMKLWCAVGNLSGGKTR--DGGSVVGASVFYKDVAGLDTEG 371

QY 60 DGEPKSAHTSPE-----KKAKELPEMDATSSRWILSTLTTSKVVIIDANQPGTVVD 113  
Db 372 SKQRSASQSSLDKLDQELKEQKFKNQEESSQVWICTSTHSTTKVVIIDAVQPGNILD 431

QY 114 QFTVCNAHVLCISSIPAAASDSYPPGEMFLDSD-----VNPEDPGADGVLAGIT 162  
Db 432 SFTVCNSHVLCIASVPGARETDYPAGEELSESGQVDKASLCGSMTSNSSAETDSLGGIT 491

QY 163 LVGC-ATRCNVPRNSCSSRGDTPVLDKGQGEVATANGKVNPSQST-EEATEATEVPDPG 220  
Db 492 VVGCSAEGVTGAATSPSTNGASPVMDKPPMEA--ENSEVDENVPTAEATEATE-ENAG 548

QY 221 PSEPETATLRPGPLTEHVFTDPAPTP-----SSGPQPGSENGPEPDSSSTRP-EPEPSGD 274  
Db 549 SAEDTVDISQTVYTEHVFTDPLGVQIPEDLSPVYQSSNDSDAYKDQISVLPNEQDLVRE 608

QY 275 PTGAGSSAAPTMMWLGQNGWLYVHSAVANWKKCLHSIKLKDSVLSLVHVKGRLVALADG 334  
Db 609 EAQKMSLLPTMWLGQNGCLYVHSSVAQWRKCLHSIKLKDSILSVHVKGIVLVALADG 668

QY 335 TLAIHFRGEDGQWDLNSYHLMOLGHPHHSIRCMVYDVRWCYKKNKVHVIOPKTMQIEK 394  
Db 669 TLAIHFRGVDGQWDLNSYHLLDLGRPHHSIRCMTVVHDKWCGYRNKIYVVPQKAMKIEK 728

QY 395 SFDAPRRRESQVRQLAWIGDGVVWSIRLDSSTLRLYHAHTHQHLQDVVDIEPYVSKMLGTGK 454  
Db 729 SFDAPRKESQVRQLAWVGDVVWSIRLDSSTLRLYHAHTYQHLQDVVDIEPYVSKMLGTGK 788

QY 455 LGFSFVRITALLVAGSRLWVGTVGNGVVISIPLTETVVLHR 494  
Db 789 LGFSFVRITALLMVSCNRLWVGTVGNGVVISIPLTESKYIFR 828

Search completed: August 23, 2004, 11:11:35  
Job time : 65 secs

---



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 11:11:38 ; Search time 55 Seconds  
(without alignments)  
3719.637 Million cell updates/sec

Title: US-10-019-495-9  
Perfect score: 3473  
Sequence: 1 MKNVPVPCRPVLEKDP...VLSKAERSHIIVQVSYTPE 651

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1295152 seqs, 314255058 residues

Total number of hits satisfying chosen parameters: 1295152

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description         |
|------------|--------|-------------|--------|-------|---------------------|
| 1          | 1732.5 | 49.9        | 1277   | 12    | US-10-276-774-2154  |
| 2          | 807.5  | 23.3        | 235    | 9     | US-09-925-297-626   |
| 3          | 511    | 14.7        | 118    | 15    | US-10-264-237-2212  |
| 4          | 460.5  | 13.3        | 766    | 14    | US-10-205-219-157   |
| 5          | 442.5  | 12.7        | 2063   | 12    | US-09-918-715-204   |
| 6          | 333.5  | 9.6         | 1240   | 12    | US-10-114-270-62    |
| 7          | 333.5  | 9.6         | 1284   | 16    | US-10-408-765A-1008 |
| 8          | 305.5  | 8.8         | 1121   | 16    | US-10-408-765A-865  |
| 9          | 304.5  | 8.8         | 1121   | 16    | US-10-408-765A-2438 |
| 10         | 186.5  | 5.4         | 119    | 15    | US-10-264-237-1721  |
| 11         | 185.5  | 5.3         | 52     | 9     | US-09-864-761-44323 |
| 12         | 161    | 4.6         | 1809   | 12    | US-10-114-270-98    |
| 13         | 161    | 4.6         | 1829   | 12    | US-10-312-352-16    |
| 14         | 154.5  | 4.4         | 1228   | 10    | US-09-917-384-1     |
| 15         | 154.5  | 4.4         | 1228   | 10    | US-09-917-383-1     |

|    |       |     |      |    |                      |                    |
|----|-------|-----|------|----|----------------------|--------------------|
| 16 | 148.5 | 4.3 | 1740 | 14 | US-10-192-381-40     | Sequence 40, Appl  |
| 17 | 146.5 | 4.2 | 985  | 10 | US-09-994-064-6      | Sequence 6, Appl   |
| 18 | 146.5 | 4.2 | 985  | 10 | US-09-994-064-66     | Sequence 66, Appl  |
| 19 | 142   | 4.1 | 3551 | 12 | US-10-263-929-144    | Sequence 144, App  |
| 20 | 141.5 | 4.1 | 582  | 14 | US-10-156-761-11938  | Sequence 11938, A  |
| 21 | 140   | 4.0 | 668  | 16 | US-10-437-963-134726 | Sequence 134726, A |
| 22 | 138   | 4.0 | 1367 | 9  | US-09-801-368-108    | Sequence 108, App  |
| 23 | 138   | 4.0 | 1538 | 16 | US-10-437-963-144952 | Sequence 144952, A |
| 24 | 135.5 | 3.9 | 112  | 15 | US-10-264-237-2046   | Sequence 2046, Ap  |
| 25 | 134   | 3.9 | 406  | 7  | US-08-523-004-2      | Sequence 2, Appl   |
| 26 | 134   | 3.9 | 406  | 8  | US-08-875-849C-2     | Sequence 2, Appl   |
| 27 | 134   | 3.9 | 406  | 9  | US-09-859-214-68     | Sequence 68, Appl  |
| 28 | 134   | 3.9 | 406  | 10 | US-09-862-989-2      | Sequence 2, Appl   |
| 29 | 134   | 3.9 | 406  | 13 | US-10-047-542-68     | Sequence 68, Appl  |
| 30 | 134   | 3.9 | 965  | 16 | US-10-437-963-190305 | Sequence 190305, A |
| 31 | 133.5 | 3.8 | 625  | 16 | US-10-437-963-165015 | Sequence 165015, A |
| 32 | 133   | 3.8 | 382  | 7  | US-08-523-004-4      | Sequence 4, Appl   |
| 33 | 133   | 3.8 | 382  | 8  | US-08-875-849C-4     | Sequence 4, Appl   |
| 34 | 133   | 3.8 | 382  | 9  | US-09-859-214-70     | Sequence 70, Appl  |
| 35 | 133   | 3.8 | 382  | 10 | US-09-862-989-4      | Sequence 4, Appl   |
| 36 | 133   | 3.8 | 731  | 13 | US-10-086-464-8      | Sequence 8, Appl   |
| 37 | 132.5 | 3.8 | 381  | 14 | US-10-165-603-20     | Sequence 20, Appl  |
| 38 | 132.5 | 3.8 | 1747 | 16 | US-10-437-963-168997 | Sequence 168997, A |
| 39 | 132   | 3.8 | 591  | 12 | US-10-425-114-68271  | Sequence 68271, A  |
| 40 | 132   | 3.8 | 1151 | 12 | US-09-825-751A-79    | Sequence 79, Appl  |
| 41 | 131.5 | 3.8 | 1555 | 16 | US-10-437-963-129858 | Sequence 129858, A |
| 42 | 131.5 | 3.8 | 1974 | 16 | US-10-647-196-24     | Sequence 24, Appl  |
| 43 | 131   | 3.8 | 584  | 14 | US-10-156-761-12405  | Sequence 12405, A  |
| 44 | 131   | 3.8 | 1843 | 16 | US-10-437-963-194983 | Sequence 194983, A |
| 45 | 130.5 | 3.8 | 297  | 16 | US-10-437-963-168840 | Sequence 168840, A |

ALIGNMENTS

RESULT 1  
US-10-276-774-2154  
; Sequence 2154, Application US/10276774  
; Publication No. US20040053245A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; APPLICANT: Tang, Y, Tom et al  
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides  
; FILE REFERENCE: 21272-030  
; CURRENT APPLICATION NUMBER: US/10/276,774  
; CURRENT FILING DATE: 2002-11-18  
; PRIOR APPLICATION NUMBER: 09/560,875  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 09/496,914  
; PRIOR FILING DATE: 2000-02-03  
; NUMBER OF SEQ ID NOS: 2700  
; SOFTWARE: Custom  
; SEQ ID NO 2154  
; LENGTH: 1277  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-276-774-2154

|                       |       |   |  |                                     |
|-----------------------|-------|---|--|-------------------------------------|
| Query Match           | 49.9% | Score 1732.5;   | DB 12;                                   | Length 1277;                        |
| Best Local Similarity | 57.3% | Pred. No. 1.1e-127;   |  |                                     |
| Matches               | 358;  | Conservative  | 79;                                      | Mismatches 141; Indels 47; Gaps 13; |
| QY                    | 1     | MKNVPVPCRPVLEKDP  | TMKLCAGVNLSGWRNEDDAGNVKFPAP-GRDPLTCDREG  | 59                                  |
| Db                    | 664   | MKNLPVPLRLPDKD  | TSMKLWCAGVNLSGGKTR--DGGSVVGASVFKDVAGLDTG | 721                                 |
| QY                    | 60    | DGEPSAHTSPE-----  | KKAKELPEMDATSSRWILTSTLTTSKVVIIDANQPGTVVD | 113                                 |
| Db                    | 722   | SKQRSASQSSSLDKLDQELKEQKELKNQEELSSLWICTSTHSATKVIIDAVQPGNILD  | 781                                      |                                     |
| QY                    | 114   | QFTVCNAHVLCISIPAAASDSYPPGEMFLDSD-----                       | VNPEDPGADGVLAGIT                         | 162                                 |
| Db                    | 782   | SFTVCNSHVLCIASVPGARETDYPAGEDLSESGQVDKASLCGSMTSSNSAETDSLGGIT | 841                                      |                                     |







```
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 865
; LENGTH: 1121
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-865

Query Match      8.8%; Score 305.5; DB 16; Length 1121;
Best Local Similarity 21.1%; Pred. No. 7.7e-15;
Matches 134; Conservative 81; Mismatches 192; Indels 227; Gaps 22;

QY 29 NLSGWRPNEDDAGNGVKPAP-----GRDPLTCDREGDGEPKSAHTSPEKKAKE-LPEMD 82
Db 610 NHMGWFCVEDD-GNHIKKEKHPLLVGHMPVMVAKQEFKIECAAYNPEPYLNNESQPDSE 668
QY 83 ATS-SRVWILTSTLTTSKVVID-ANQPGTVVDQFTVCNAHVLCISSIPAASDSDYPPGE 140
Db 669 STAHGFLWIGSCTHMQGQIAIVSFQNSTPKVIECFNV-ESRILCMLYVPV----- 717
QY 141 MFLDSVNPEDPGADGVLGITLVGCATRCNVPRSNCSRRGDTPVLDKQGGEVATIANGK 200
Db 718 -----EEXRREPGA----- 726
QY 201 VNPSQSTEATEATEVPDPGPSEPETATLRPGPLTEHVFTDPAPTSSGPGPSENGPEP 260
Db 727 -----PPDPETPAVR----- 736
QY 261 DSSSTRPEPEPSGDPGTGAGSSAAPTMLGAQNGWLYVHSAVANWKKLHSLIKL----- 313
Db 737 -----ASDVPTICVGTGEGSISYKSSQGSKK-----VRLQHFFFTPE 773
QY 314 KDSVLSLVHVKGRVLVALADGTLAIFHRGEDGQWDLNHYHLMDLGHPHHSIRCMVVYDR 373
Db 774 KSTVMSLACTSQSLYAGLVNGAVASYARAPDGSWDSEPOKVIKLG--VLPVRSLLMMEDT 831
QY 374 VMCYKKNKVHVIQPKTMQIEKSFDAHPRRESQVRQLAWIGDGVVWSIRLDSLRLYHAHT 433
Db 832 LWAASGGQVFIISVETHAVEGQLEAHQEGMVISHMAVSGVGIIWIAFTSGSTLRLFHTET 891
QY 434 HQHLQDVDIPEYVSKMLGTGKLGFSFVRITALLVAGSRLMVGTGNGVVISIPLTETVVLH 493
Db 892 LKHLQDINIATPVHNMLP-----GHQRLSVTSLLVCHGLMVGTSLGLVALPV----- 940
QY 494 RGQLLGLFRANKTSPTS GEGARPGGIIHVYGDSSDRAASSFIPYCSMAQAQLCFHGHDA 553
Db 941 -----PRLQGIKVTG-----RGMVSYHAHNSP 963
QY 554 VKFFVSPGNVLATL-----NGSVLDSPAEGPGPAAPASEVEGOKLRNVLVLSGGEYID 608
Db 964 VKFI-----VLATALHEKDKKSRDSLAPGPEP-----QDEDQK---DALPSGGAG--- 1006
QY 609 FRIGDGEDDET---EEGAGDMSQVKPVLKAERS 639
Db 1007 SSLSQGDPDAAIWLGDSLGSMTQ-KSDLSSSSGS 1039
```

```
RESULT 9
US-10-408-765A-2438
; Sequence 2438, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
```

```
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2438
; LENGTH: 1121
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2438

Query Match      8.8%; Score 304.5; DB 16; Length 1121;
Best Local Similarity 21.1%; Pred. No. 9.3e-15;
Matches 134; Conservative 81; Mismatches 192; Indels 227; Gaps 22;

QY 29 NLSGWRPNEDDAGNGVKPAP-----GRDPLTCDREGDGEPKSAHTSPEKKAKE-LPEMD 82
Db 610 NHMGWFCVEDD-GNHIKKEKHPLLVGHMPVMVAKQEFKIECAAYNPEPYLNNESQPDSE 668
QY 83 ATS-SRVWILTSTLTTSKVVID-ANQPGTVVDQFTVCNAHVLCISSIPAASDSDYPPGE 140
Db 669 STAHGFLWIGSCTHMQGQIAIVSFQNSTPKVIECFNV-ESRILCMLYVPV----- 717
QY 141 MFLDSVNPEDPGADGVLGITLVGCATRCNVPRSNCSRRGDTPVLDKQGGEVATIANGK 200
Db 718 -----EEXRREPGA----- 726
QY 201 VNPSQSTEATEATEVPDPGPSEPETATLRPGPLTEHVFTDPAPTSSGPGPSENGPEP 260
Db 727 -----PPDPETPAVR----- 736
QY 261 DSSSTRPEPEPSGDPGTGAGSSAAPTMLGAQNGWLYVHSAVANWKKLHSLIKL----- 313
Db 737 -----ASDVPTICVGTGEGSISYKSSQGSKK-----VRLQHFFFTPE 773
QY 314 KDSVLSLVHVKGRVLVALADGTLAIFHRGEDGQWDLNHYHLMDLGHPHHSIRCMVVYDR 373
Db 774 KSTVMSLACTSQSLYAGLVNGAVASYARAPDGSWDSEPOKVIKLG--VLPVRSLLMMEDT 831
QY 374 VMCYKKNKVHVIQPKTMQIEKSFDAHPRRESQVRQLAWIGDGVVWSIRLDSLRLYHAHT 433
Db 832 LWAASGGQVFIISVETHAVEGQLEAHQEGMVISHMAVSGVGIIWIAFTSGSTLRLFHTET 891
QY 434 HQHLQDVDIPEYVSKMLGTGKLGFSFVRITALLVAGSRLMVGTGNGVVISIPLTETVVLH 493
Db 892 LKHLQDINIATPVHNMLP-----GHQRLSVTSLLVCHGLMVGTSLGLVALPV----- 940
QY 494 RGQLLGLFRANKTSPTS GEGARPGGIIHVYGDSSDRAASSFIPYCSMAQAQLCFHGHDA 553
Db 941 -----PRLQGIKVTG-----RGMVSYHAHNSP 963
QY 554 VKFFVSPGNVLATL-----NGSVLDSPAEGPGPAAPASEVEGOKLRNVLVLSGGEYID 608
Db 964 VKFI-----VLATALHEKDKKSRDSLAPGPEP-----QDEDQK---DALPSGGAG--- 1006
QY 609 FRIGDGEDDET---EEGAGDMSQVKPVLKAERS 639
Db 1007 SSLSQGDPDAAIWLGDSLGSMTQ-KSDLSSSSGS 1039
```

```
RESULT 10
US-10-264-237-1721
; Sequence 1721, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA131PI
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
```

```
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 1721
; LENGTH: 119
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (108)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (116)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-237-1721

Query Match      5.4%; Score 186.5; DB 15; Length 119;
Best Local Similarity 37.7%; Pred. No. 8.3e-07;
Matches 46; Conservative 22; Mismatches 43; Indels 11; Gaps 3;

QY 367 MAVVYDRVWCYKKNVHVIOPTKMQIEKSFDAHPRRESQVRQLAWIGD---GVWVSIRLD 423
Db 1 MVSVGRLWCQCQNRVLVLSPTDLQLEHMFYV---GQDSSRCVACMVDSSLGWVTLKGS 57

QY 424 STLRLYHAHTHQLDVDIEPVYSKMLGTG-----KLGFSFVRITALLVAGSRLLWVGTDN 478
Db 58 AHVCLYHPDPTFEQLAEVDVTPPVHRMLASSDAIRQHKAACLRITALLVCXELLWVGTXA 117

QY 479 GV 480
Db 118 GV 119

RESULT 11
US-09-864-761-44323
; Sequence 44323, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 44323
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005920.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: SWISSPROT HIT: P34609, EVALUE 3.00e-09
; OTHER INFORMATION: EST_HUMAN HIT: BF131060.1, EVALUE 6.00e-26
US-09-864-761-44323

Query Match      5.3%; Score 185.5; DB 9; Length 52;
Best Local Similarity 66.7%; Pred. No. 3.1e-07;
Matches 36; Conservative 5; Mismatches 10; Indels 3; Gaps 2;
```

```
QY 503 NKTSPSTSGEGARPGGIHVYGDSSDRAA-SSFIPYCSMAQAQLCFHGHDAVK 555
Db 1 NKTSGVPGN--RPGSVIRVYGDENSDKVTPTGTFIPYCSMAHAQAQLCFHGHDAVK 52

RESULT 12
US-10-114-270-98
; Sequence 98, Application US/10114270
; Publication No. US20040030110A1
; GENERAL INFORMATION:
; APPLICANT: Guo, Xiaojia
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Miller, Charles E.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Patturajan, Meera
; APPLICANT: Liu, Ziaohong
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Li, Li
; APPLICANT: Vernet, Corine
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Gorman, Linda
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Smithson, Glennda
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Taupier Jr., Raymond J.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Ji, Weizhen
; APPLICANT: Anderson, David W.
; APPLICANT: Liete, Mario W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Stone, David J.
; APPLICANT: MacDougall, John R.
```

; APPLICANT: Rothenberg, Mark E.  
; TITLE OF INVENTION: No. US20040030110A1e1 Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-322C  
; CURRENT APPLICATION NUMBER: US/10/114,270  
; CURRENT FILING DATE: 2002-11-27  
; PRIOR APPLICATION NUMBER: 60/281,086  
; PRIOR FILING DATE: 2001-04-03  
; PRIOR APPLICATION NUMBER: 60/281,136  
; PRIOR FILING DATE: 2001-04-03  
; PRIOR APPLICATION NUMBER: 60/281,863  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/281,906  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/282,020  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: 60/282,930  
; PRIOR FILING DATE: 2001-04-10  
; PRIOR APPLICATION NUMBER: 60/282,934  
; PRIOR FILING DATE: 2001-04-10  
; PRIOR APPLICATION NUMBER: 60/283,512  
; PRIOR FILING DATE: 2001-04-12  
; PRIOR APPLICATION NUMBER: 60/283,710  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: 60/284,234  
; PRIOR FILING DATE: 2001-04-17  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 470  
; SEQ ID NO 98  
; LENGTH: 1809  
; TYPE: PRP  
; ORGANISM: Homo sapiens  
US-10-114-270-98

Query Match 4.6%; Score 161; DB 12; Length 1809;  
Best Local Similarity 20.7%; Pred. No. 0.0041;  
Matches 138; Conservative 86; Mismatches 290; Indels 152; Gaps 28;  
QY 37 EDDAGNVKPA-----PGRDPLTCDREGDGPKSAHTSPEKKKAKELPEMDATSSR---- 87  
Db 1228 DPERGSLASPAFSPRSPAWIPVPARREAEKVPREERKSPEDKSMILSVLDTSLQRPAGL 1287  
QY 88 --VWILTSTLTSKVVIIDANQPGTVVDQFTVCNAHVLCISSIPAASDSDYPPGEMFLDS 145  
Db 1288 IWVHATSNQEP SRLGGAEEERPGT--PELAPAPMQSAVAE-PLPSPRAQPPGG--TPA 1342  
QY 146 DVNP-----EDPGADGVLAGITLVGCATRCNVPRSNCSRGDTPVLDKQGEVATIANG 199  
Db 1343 DAGPGQGSSEEP--ELVFA-----VNLPPAQLSSD-----EETREELARI--G 1383  
QY 200 KVNPSQS--TEEATEATEVDPDPSEPETATLRPGPLTHEVFTDPAPTPSSGPQPSGENCEP 258  
Db 1384 LVPPPEEFANGVLLATPLAGPGPS-PTTV---PSPASGKESSEPPAPESAADSGVE--- 1436  
QY 259 EPDSSSTRPEPEPSGDPPTGAGSSAAPTMMWLGANQWLY-VHSAVANWKKLHSLKLDVS 317  
Db 1437 ---EADTRSSDPHLETTSTISTVSSMSTLSSESGELTDTHTSFADG-----HTFLLEKPP 1489  
QY 318 LSLVHVKGKRVLVALADGTILA---IFHRGEGQWDLNSVHMLDGLPHPHSIRCMAVVYDR 373  
Db 1490 ---VPPKPKLKSPLGKGPVTFRDPLLLKQSSDSELMAQQHHAASAG----- 1531  
QY 374 VWCYKKNKVHVQPKTMQIEKSFDAHPRESQVRLANIGDGV-----WVSI 420  
Db 1532 -----LASAAGPARPRVLFQRRSKLW-GDPVESRGLPGPEDDKPTVIS 1573  
QY 421 RLDSTLRLYHAHTHQHLQDQVD-----JEPYVSKMLGTGKLGFSFVRITALLVAGSRLWV 474  
Db 1574 ELSSRLQQLNKDTRSLGEEFVGGGLSLDLPACKSPIAAARLFSSLGELSS--ISAQRSPG 1631  
QY 475 GTGNGVVVISIPLTETVVLHKGQLLGLRANKTSPTSGEGARPGGIHVVGGDDSSDRAASSF 534  
Db 1632 GPGGGASISVRPSGRYPVARRAPSPVKPASLERVEGLGAGAGGAGRPFG-----L 1681

QY 535 IPYCSMAQAQLCFHGHRDAVKFFVSVPGNVLATLNGSVLDSPAEGPGPAAPASEVEGQKL 594  
Db 1682 TPPTILKSSLSIPHEPKKEVRFVRSVSARSRSPPSPPLSPSPASGPGPGAPGRRPFQOK 1741  
QY 595 RNVLVLSGGEGYIDFRIGD-----GE-----DDETEEGAGDMSQVKPVLSKAERSHII 642  
Db 1742 PLQL-----WSKFDVGDWLESIHGHRDRFEDHEIEGAH-----LPALTCKDDFVELG 1789  
QY 643 VMQVSY 648  
Db 1790 VTRVGH 1795

RESULT 13

US-10-312-352-16  
; Sequence 16, Application US/10312352  
; Publication No. US20040053824A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.; TANG, Y. Tom  
; APPLICANT: YUE, Henry; AZIMZAI, Yalda  
; APPLICANT: HE, Ann; BATRA, Sajeev  
; APPLICANT: LO, Terence P.; NGUYEN, Danniel B.  
; APPLICANT: BURRILL, John D.; MARCUS, Gregory A.  
; APPLICANT: ZINGLER, Kurt A.; GANDHI, Ameena R.  
; APPLICANT: LAL, Preeti G.; KEARNEY, Liam  
; APPLICANT: BURFORD, Neil; YAO, Monique G.  
; APPLICANT: CHAWLA, Narinder K.; ELLIOT, Vicki S.  
; APPLICANT: ARVIZU, Chandra S.; KHAN, Farrah A.  
; APPLICANT: BAUGHN, Mariah R.; HAFALIA, April, J.A.  
; APPLICANT: POLICKY, Jennifer L.; AU-YOUNG, Janice K.  
; APPLICANT: LU, Yan; BOROWSKY, Mark L.  
; APPLICANT: LU, Dyung Aina M.; RAMKUMAR, Jayalaxmi  
; APPLICANT: YANG, Junming; GURURAJAN, Rajagopal  
; APPLICANT: WARREN, Bridget A.; GIETZEN, Kimberly J.  
; APPLICANT: XU, Yuming; KALLICK, Deborah A.  
; APPLICANT: LEE, Ernestine A.; THANGAVELU, Kavitha  
; APPLICANT: DELEGEANE, Angelo M.; LEE, Sally  
; TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES  
; FILE REFERENCE: PF-0794 USN  
; CURRENT APPLICATION NUMBER: US/10/312,352  
; CURRENT FILING DATE: 2002-12-18  
; PRIOR APPLICATION NUMBER: PCT/US01/21067  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/215,454  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 60/219,462  
; PRIOR FILING DATE: 2000-07-18  
; PRIOR APPLICATION NUMBER: US 60/240,111  
; PRIOR FILING DATE: 2000-10-12  
; PRIOR APPLICATION NUMBER: US 60/240,106  
; PRIOR FILING DATE: 2000-10-12  
; PRIOR APPLICATION NUMBER: US 60/244,021  
; PRIOR FILING DATE: 2000-10-27  
; PRIOR APPLICATION NUMBER: US 60/248,887  
; PRIOR FILING DATE: 2000-11-14  
; PRIOR APPLICATION NUMBER: US 60/249,570  
; PRIOR FILING DATE: 2000-11-16  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: PERL Program  
; SEQ ID NO 16  
; LENGTH: 1829  
; TYPE: PRP  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. US20040053824A1 7477312CD1  
US-10-312-352-16

Query Match 4.6%; Score 161; DB 12; Length 1829;  
Best Local Similarity 20.7%; Pred. No. 0.0042;  
Matches 138; Conservative 86; Mismatches 290; Indels 152; Gaps 28;  
QY 37 EDDAGNVKPA-----PGRDPLTCDREGDGPKSAHTSPEKKKAKELPEMDATSSR---- 87

Db 1248 DPERGSLASPAFSPRPAWIPVPARREAEKVPREERKSPEDKKSMILSVLDTSLORPAGL 1307  
QY 88 --VWILTSTLTTSKVVIIDANQPGTVVDQFTVCNAHVLCISSIPAAASDSDDYPPGEMFLDS 145  
Db 1308 IVVHATSNQOEPSRLGGAEEERPGT--PELAPAPMQSAVAE-PLPSPRAQPPGG--TPA 1362  
QY 146 DVNP-----EDPGADGVLGATILVGCATRCNVPRSNCSRGDTPVLDKGQGEVATIANG 199  
Db 1363 DAGPGQGSSEEB--ELVFA-----VNLPPAQLSSSD-----EETREELARI--G 1403  
QY 200 KVNPSQS--TEEATEATEVPDPGPSEPETATLRPGPLTEHVFTDPAPTSSGGPQGSNGP 258  
Db 1404 LVPPPEEFANGVLLATPLAGPGFS--PTTV---PSPASCKPSSSEPPPAPESAADSGVE--- 1456  
QY 259 EPDSSSTRPEPEPSGDPGTGAGSSAAPTMLGAQNGWLY-VHSAVANWKKCLHSIKLKDSV 317  
Db 1457 ---EADTRSSSDPHLETTSTISTVSSMSTLSSSEGLTDTHTSFADG---HTFLEKPP 1509  
QY 318 LSLVHVKGRLVALADGTIA----IFHRGEDGQWDLSNYHLMDLGHPHSIRCMAVVYDR 373  
Db 1510 ---VPPKPKLKSPLCKGPVTRFDRLILKQSSDSELMAQQHHAASAG----- 1551  
QY 374 VMCGYKNKVHVIQPKTMQIEKSFDAHPRRRESQVRLAWIGDV-----WVSI 420  
Db 1552 -----LASAAGPARPRYLFORRSKLW-GDPVBSRGLPGPEDDKPTVIS 1593  
QY 421 RLDSTLRLYHAHTHQHLQDVD-----IEPVVSKMLGTGKLGFSFVRITALLVAGSRLWV 474  
Db 1594 ELSRLQQLNKDTRSLGEEPVGGLSLDLPKKSPIAARLFSSLGELSS--ISAQRSPG 1651  
QY 475 GTGNGVVISIPLTETVVLHRGQLGLRANKTSPTSGEGARPGGIHVYGDSSDRAASSF 534  
Db 1652 GPGGGASYSVRPSGRYPVARRAPSPVKPASLERVEGLGAGAGGAPPG-----L 1701  
QY 535 IPYCSMAQAQLCFHGHRDAVKFFVSPGNVLATNGSVLDSPAEGPGPAAPASEVEGQKL 594  
Db 1702 TPPTILKSSLSIPHEPKKEVRFVRSVSARSRSRSPSPPLPSPASGPGPGAGPRRPFQOK 1761  
QY 595 RNVVLVSGEGYIDFRIGD-----GE-----DDETEEGAGDMSQVKPVLKSAERSHII 642  
Db 1762 PLQL-----WSKPDVGDWLESIHLGEHRDRFEDHIEGAH-----LPALTCKDDFVELG 1809  
QY 643 VMQVSY 648  
Db 1810 VTRVGH 1815

RESULT 14  
US-09-917-384-1  
; Sequence 1, Application US/09917384  
; Publication No. US20030096342A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,384  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1228  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-384-1

Query Match 4.4%; Score 154.5; DB 10; Length 1228;  
Best Local Similarity 23.6%; Pred. No. 0.0078;  
Matches 81; Conservative 27; Mismatches 112; Indels 123; Gaps 16;  
QY 16 KDPTMKLWCAAG-----VNLSGWRPNEDDAGNGV-----KPAPGRDPLTCD--REG 59  
Db 288 KEFFMTATQQVGGQPVESANFYQWNPDIIDEADYAVDLYSRLVAAGFPSSIGMLIDLRLNG 347  
QY 60 DGEPKSAHTSPEKKKAKELPEMDATSSRVWILTSTLTTSKVVIIDANQPGTVVDQFTV-- 117  
Db 348 WGGPNEP-TGP-----STAT-----DVN---TFVNQSKIDL 374  
QY 118 -----CNAHVLCISSIPAAASDSDYPPGEMFLDSDVNPEDPG-ADGVLAG----- 160  
Db 375 RQHRGLWCNQNGAGLGQPPQASPTDFP--NAHLDAYVVIKPPGESDGTSAASDPTTGKS 432  
QY 161 -----ITLVGCAT-----RCNVPRSNCSRGDTPVLDK 188  
Db 433 DMCDDPTTTSYGVLTNALPNSFIAGQWFPAQFDQLVANARPAVPTSTSSPPPP-- 488  
QY 189 GQGEVATIANGKVNPSQSTEEATEATEVPDPGPSEPETATLRPGPLTEHVFTDPAPTSS 248  
Db 489 -----PSPASPSPPSPSSSSPSPSPSSSSPSPSPSPSPSPSPSPSPSPSPSPSPSS 537  
QY 249 GPQPGSENGPEPDSS-----STRPEPEPSGDPGTGAGS-SAAPT 285  
Db 538 SPSPSPSPSPSPSSSPSPSPSSSPSPSPSPSPSPSPSSSPSPSPSPSPSPT 580

RESULT 15  
US-09-917-383-1  
; Sequence 1, Application US/09917383  
; Publication No. US20030104522A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,383  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1228  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-383-1

Query Match 4.4%; Score 154.5; DB 10; Length 1228;  
Best Local Similarity 23.6%; Pred. No. 0.0078;  
Matches 81; Conservative 27; Mismatches 112; Indels 123; Gaps 16;  
QY 16 KDPTMKLWCAAG-----VNLSGWRPNEDDAGNGV-----KPAPGRDPLTCD--REG 59  
Db 288 KEFFMTATQQVGGQPVESANFYQWNPDIIDEADYAVDLYSRLVAAGFPSSIGMLIDLRLNG 347  
QY 60 DGEPKSAHTSPEKKKAKELPEMDATSSRVWILTSTLTTSKVVIIDANQPGTVVDQFTV-- 117  
Db 348 WGGPNEP-TGP-----STAT-----DVN---TFVNQSKIDL 374  
QY 118 -----CNAHVLCISSIPAAASDSDYPPGEMFLDSDVNPEDPG-ADGVLAG----- 160  
Db 375 RQHRGLWCNQNGAGLGQPPQASPTDFP--NAHLDAYVVIKPPGESDGTSAASDPTTGKS 432  
QY 161 -----ITLVGCAT-----RCNVPRSNCSRGDTPVLDK 188

|    |     |           |        |         |        |        |                |          |           |              |
|----|-----|-----------|--------|---------|--------|--------|----------------|----------|-----------|--------------|
| Db | 433 | DEMCDPYTT | SYGVL  | TNALP   | NSPIAG | WFPAQ  | FDQLVANARPAVPT | STSSPPPP | ----      | 488          |
| QY | 189 | GQGEVATI  | ANGKVN | PSQSTEE | ATEATE | VPDPG  | SEPE           | TATLRPG  | ELTEHVFTD | PAPTPSS 248  |
| Db | 489 | -----     | PSPGAS | PSPPSP  | SSSPSP | SSSPSP | SSSPSP         | SSSPSP   | SSSPSP    | SSSPSPSS 537 |
| QY | 249 | GPQPGSE   | NGPEPD | SS----- | STRPE  | PEPSG  | DPTGAGS        | -SAAPT   |           | 285          |
| Db | 538 | SPSPSP    | SPSPSP | SPSPSP  | SPSPSP | SPSPSP | SPSPSP         | SPSPSP   | SPSPSP    | SPSPSPT 580  |

Search completed: August 23, 2004, 11:16:35  
Job time : 59 secs









RESULT 8  
S48478  
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)  
N;Alternate names: extracellular glucoamylase; mucin-like protein MUC1; protein YIR019c  
C;Species: Saccharomyces cerevisiae  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 12-Nov-1999  
C;Accession: S48478; A26877; B26877; S27281; JC6123  
R;Rowley, K.  
submitted to the EMBL Data Library, October 1994  
A;Reference number: S48478  
A;Accession: S48478  
A;Molecule type: DNA  
A;Residues: 1-1367 <ROW>  
A;Cross-references: GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364; GSPDB:GNO0009; MIP  
R;Yamashita, I.; Nakamura, M.; Fukui, S.  
J. Bacteriol. 169, 2142-2149, 1987  
A;Title: Gene fusion is a possible mechanism underlying the evolution of STA1.  
A;Reference number: A91831; MUID:87194600; PMID:3106330  
A;Accession: A26877  
A;Molecule type: DNA  
A;Residues: 1-242 <YAM>  
A;Cross-references: EMBL:M16164; NID:gl72522; PIDN:AAA35014.1; PID:gl72525  
A;Accession: B26877  
A;Molecule type: DNA  
A;Residues: 762-1331 <YA2>  
A;Cross-references: EMBL:M16165; NID:gl72523; PIDN:AAA35015.1; PID:gl72526  
R;Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.  
FEBS Lett. 239, 179-184, 1988  
A;Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar  
A;Reference number: S27281; MUID:89031230; PMID:3141213  
A;Accession: S27281  
A;Molecule type: DNA  
A;Residues: 1-31 <PAR>  
A;Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552  
R;Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.  
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996  
A;Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohy  
A;Reference number: JC6123; MUID:96323237; PMID:8710886  
A;Accession: JC6123  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 1-1367 <LAM>  
A;Cross-references: GB:U30626; NID:gl1304386; PIDN:AAC49609.1; PID:gl1304387  
C;Genetics:  
A;Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458  
A;Cross-references: MIPS:YIR019c; SGD:S0001458  
A;Map position: 9R  
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase  
C;Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein  
F;5-21/Domain: transmembrane #status predicted <TM1>  
F;1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 4.0%; Score 138; DB 1; Length 1367;  
Best Local Similarity 24.2%; Pred. No. 0.41;  
Matches 60; Conservative 25; Mismatches 121; Indels 42; Gaps 8;

QY 47 APGRDPLTCDREGDGEPKSAHTSPEKKKAKELPEMDATSSRVWILTLTTSKVVIIDAN 106  
Db 456 APVPTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPVPTP 515  
QY 107 QPGTVVDQFTVCNAHVLCISSIPAAASDSYPPGEMFLDSVNPEDPGADGVLGITLVGC 166  
Db 516 SSSTTES-----SSAPAPT-----PSSSTTESSAP-----VTSSTTES 550  
QY 167 ATRCNVPRNSCSRGDTPV-----LDKGQGEVATIANGKVNPSQSTEBATEATEVPDPGPS 222  
Db 551 SAPVPTPSSSTTESSTPTVTSSTTESSAPVPT-----PSSSTTESSA-PVPTPSSS 602  
QY 223 EPETAITL-RPGPLTEHVFTDPAPTPSSGPGQSGENGPEPDSSSTR-----PEPEPSGDP 277  
Db 603 TTESSSAPAPTTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTE 662  
QY 278 AGSSAAAPT 285

Db 663 SSSAPVPT 670  
RESULT 9  
T29018  
hypothetical protein ZK84.1 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Mar-2000  
C;Accession: T29018  
R;Kirsten, J.  
submitted to the EMBL Data Library, April 1995  
A;Description: The sequence of C. elegans cosmid ZK84.  
A;Reference number: Z20553  
A;Accession: T29018  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-801 <KIR>  
A;Cross-references: EMBL:U23181; PIDN:AAC48204.1; GSPDB:GNO0020; CESP:ZK84.1  
A;Experimental source: strain Bristol N2; clone ZK84  
C;Genetics:  
A;Gene: CESP:ZK84.1  
A;Map position: 2  
A;Introns: 22/2; 45/3; 108/1  
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
Query Match 3.9%; Score 135; DB 2; Length 801;  
Best Local Similarity 22.6%; Pred. No. 0.32;  
Matches 72; Conservative 37; Mismatches 143; Indels 66; Gaps 15;

QY 2 KNPVPVVCYCRPLVEKDPTMKLMCAAGVNLGWRPNEDDAGN--GVKPAPGRD-----PL 53  
Db 282 EEAPAPA---PTAETPAPETVSAAPEAANSY-----DSAGGDAATAPAPSEAAADAAAPT 334  
QY 54 TCDREGDGEPKSAHTSPEKKKAK--ELPEMDATSSRVWILTLTTSKVVIIDA-----NQ 107  
Db 335 DSAASADTTAALVDVTSSEHAAESTEAPATDIAATE----TTPAPSVPAFVADAAAGYDS 390  
QY 108 PGTVVDQFTVCNAHVLCISSIPAAASDSYPPG----EMFLDSVNPEDPGADGVLGITL 163  
Db 391 PSSIPEE-----TPAPAAEDTPAPASAAAEETPAPAPAAETPAPETASAAPDA 439  
QY 164 VGCATRCNV--PRSNCSRGDTPVLDKGQGEVATIANGKVNPSQSTEEAT-----EATEV 216  
Db 440 AGGAAPADVAAPADVATTAPETSSAQSAAGSYDV-----PSEPASEVTAPIVESATEA 492  
QY 217 PD-----PGPSBPETATLRPGPLTEHVFTDPAPTPSSGPGQSGENGPEPDSSSTRPE- 268  
Db 493 PSDSAAAPIGPAASEPAPAPIE-APATDAATLETAPAPAAEPAPAPAEAAAAGYDAPSSVPEE 551  
QY 269 ---PEPSGDPTGAGSSAA 283  
Db 552 TPAPAPAADETPAPAPAA 569  
RESULT 10  
B86369  
hypothetical protein F508.10 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
C;Accession: B86369  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: B86369

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-731 <STO>  
A;Cross-references: GB:AE005172; NID:g4056437; PIDN:AAC98010.1; GSPDB:GN00141  
C;Genetics:  
A;Map position: 1

Query Match 3.8%; Score 133; DB 2; Length 731;  
Best Local Similarity 20.3%; Pred. No. 0.39;  
Matches 110; Conservative 61; Mismatches 212; Indels 158; Gaps 24;

QY 64 KSAHTSPEKKKAKELPEMDATSSRVWILTSTLTTSKVWIIDANQPGTVVDQFTVCNAHVL 123  
Db 6 ESPSSPPAPPADTAPPETPEB-----NSALPPVDSSPPSPADS----- 46

QY 124 CISSIPAAASDYPGEMFL-----DSNVNPE--PGADGVLAGITLVGCATRC 170  
Db 47 --SSTPPLSEPTPPDSQLPPLPSILPPLTDSPPPSDSSPPVDSTPS----- 93

QY 171 NVPRNSCSRGDTPVLDKQGEVATIANGKVN--PSQSTEEATEATEVDPGPSEPETAT 228  
Db 94 --PPPPTSNESPPPEDSETPPAPPNESNDNNPPPSQDLQSPPPSPSPNVGPTNPESPP 151

QY 229 LR--PGPLTEHV-----TDPAPTSSGP---QPGSENGPEPDSSSTRPEPEPSG 273  
Db 152 LQSPAPPASDPTNSPPASPLDPTNPPIQSPGATSPPANPNAP-PSPPFTVPPKTPSS 210

QY 274 DPTGAGSSAAPTMWL-----GAQNGWLYVHSAVANWKKCLHSIKLKDSVLS 319  
Db 211 GPVVSPLTSFSKGTPTPNQNGDGGGGGYQGKTMVGMAVAGF-----AIMA 259

QY 320 LVHVKGRVLVALADGTLAIFHRGEDGQWDLNHYHLMDLGHPHHSIRCMVYVDR-VWCGY 378  
Db 260 LI-----GVVFLVRKKRNIDSYN-HSQYLPHNFVSVKSDGFLYGQDPKGKY 306

QY 379 KNKVHVIQKTMQIEKSFDAHPRRESQVRLAWIGDVVWSIRLSTLRLYHAHTHQLQ 438  
Db 307 SSGPNGSMYNNSSQQQS-----SMGNSYGTAGG-----YPHHQMQSSG 345

QY 439 DVDIEPVYSKMLGTGKLGFSFVRITALLVAGSR---LWVGNGVWVISIPLTETVVLHRG 495  
Db 346 TPD-----SAILGSGQTHFSYEELAEITQGFARKNILGEG-GFGCVYKGLQD-----G 393

QY 496 QLLGLRANKTSPTSGEGARP-----GGIHHVYDDSDRAASSFIPYCSMAQAQLCF 547  
Db 394 KWAVKQLKAG--SQQGDREFKAEVEIISRHH-----RHLVSLVGYCISDQHRLLI 443

QY 548 H 548  
Db 444 Y 444

RESULT 11  
T09219  
basal transcription factor SNAPc large chain SNAP190 - human  
C;Species: Homo sapiens (man)  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C;Accession: T09219  
R;Wong, M.W.; Henry, R.W.; Ma, B.; Kobayashi, R.; Klages, N.; Matthias, P.; Strubin, M.; Mol. Cell. Biol. 18, 368-377, 1998  
A;Title: The large subunit of basal transcription factor SNAPc is a Myb domain protein b  
A;Reference number: Z16616; MUID:98078693; PMID:9418884  
A;Accession: T09219  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-1469 <WON>  
A;Cross-references: EMBL:AF032387; NID:g2641556; PIDN:AAC02972.1; PID:g2641557  
A;Experimental source: tissue type fetal cell teratocarcinoma  
C;Genetics:  
A;Gene: SNAP190  
C;Function:  
A;Description: transcription factor; required for transcription of snRNA genes  
C;Superfamily: unassigned myb DNA-binding repeat proteins; myb DNA-binding repeat homol

C;Keywords: DNA binding; RNA biosynthesis; transcription factor  
F;396-447/Domain: myb DNA-binding repeat homology <MYB>

Query Match 3.8%; Score 133; DB 2; Length 1469;  
Best Local Similarity 22.9%; Pred. No. 0.97;  
Matches 88; Conservative 28; Mismatches 133; Indels 136; Gaps 19;

QY 3 NVPVPVYCRPLVEKDPMTKMLWCAAGVNLSGWRPNEDDA-----GNGVKPA----- 47  
Db 949 NVPLSGPGAPAAAKPGTSGSQWEAGTSAKDKKLSTMQALPLAPVFSEAGTAPAAASQAPA 1008

QY 48 --PGRDPLTCDREGDGEPEKSAHTSPEKKKAKELPEMD----- 82  
Db 1009 LGPGQISVSCPESGLGQSQ-----APAAARKQGLPEAPFFLPGAPSPPLFVQPLSLTHIG 1064

QY 83 ----ATSSRV---WILTS-----TLTTSKVVII--DANQPG-----TVVDQFTVCNAHVL 123  
Db 1065 GPHVATSVPLPVTWVLTAQGLLPVPVPAVVSLPRPAGTGGPAGLLATLLPPLTETRA--A 1122

QY 124 CISSIPAAASDYPGEMFLDSVDNPEDPGADGVLAGITLVGCATRCNVPRNSCSRGD 183  
Db 1123 QGPRAPALSSSWQPPANM-----NREPEP-----SCRTDTPAPPTHALSQSPA 1165

QY 184 PVLDKQGEVATIANGKVNPSQSTEEATEATEV-----DPGPSEPETATLRPG--- 232  
Db 1166 ----EADGSVAFPVG-----EAQVAREIPEPRTSSHADPPEAEPPWSGRLPAFGG 1211

QY 233 --PLTEHVETDPAPTSS-----GP-----QPGSENGPEPDSSSTRPEPEPSG 273  
Db 1212 VIPATE-----PRGTPGSPSGTQEPGRGLGLEKPLRLRQPGKEGALDLEKPLPQGPPEK 1266

QY 274 DPTGAG-----SSAAPTMWLGAQNG 293  
Db 1267 GALDLGLLSQEGEAATQWLGGQRG 1291

RESULT 12  
IUMSNL

neural cell adhesion molecule 1 precursor, long domain splice form - mouse  
N;Alternate names: NCAM-180

N;Contains: neural cell adhesion molecule, short domain splice form (NCAM-140)  
C;Species: Mus musculus (house mouse)

C;Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 31-Dec-2000  
C;Accession: A29673; S00844; S00384; A28281; A44290; S00383

R;Barthels, D.; Santoni, M.J.; Wille, W.; Ruppert, C.; Chaix, J.C.; Hirsch, M.R.; Fontec:  
EMBO J. 6, 907-914, 1987

A;Title: Isolation and nucleotide sequence of mouse NCAM cDNA that codes for a Mr 79,000  
A;Reference number: A29673; MUID:87246524; PMID:3595563

A;Accession: A29673  
A;Molecule type: mRNA

A;Residues: 1-548,'T',550-571,'T',573-574,'D',576-588,'MQPS',593,'S',595-599,'P',601,'L',  
A;Cross-references: EMBL:Y00051; NID:G53342; PIDN:CAA68263.1; PID:G53343

R;Santoni, M.J.; Barthels, D.; Barbas, J.A.; Hirsch, M.R.; Steinmetz, M.; Goridis, C.; W:  
Nucleic Acids Res. 15, 8621-8641, 1987

A;Title: Analysis of cDNA clones that code for the transmembrane forms of the mouse neur  
A;Reference number: S00844; MUID:88067687; PMID:3694567

A;Accession: S00844  
A;Molecule type: mRNA

A;Residues: 529-809,1077-1115 <SAN>  
A;Cross-references: EMBL:X06328; NID:G53322; PIDN:CAA29641.1; PID:g817984

R;Barbas, J.A.; Chaix, J.C.; Steinmetz, M.; Goridis, C.  
EMBO J. 7, 625-632, 1988

A;Title: Differential splicing and alternative polyadenylation generates distinct NCAM tr  
A;Reference number: S00382; MUID:88283628; PMID:3396534

A;Accession: S00384  
A;Molecule type: DNA

A;Residues: 642-1115 <BAR>  
A;Cross-references: EMBL:X07195

R;Barthels, D.; Vopper, G.; Wille, W.  
Nucleic Acids Res. 16, 4217-4225, 1988

A;Title: NCAM-180, the large isoform of the neural cell adhesion molecule of the mouse, i  
A;Reference number: A28281; MUID:88247737; PMID:2454455

A;Accession: A28281

A;Molecule type: mRNA  
A;Residues: 804-1081 <BA3>  
A;Cross-references: EMBL:X07244; NID:g53321; PIDN:CAA30230.1; PID:g929720  
R;Rougou, G.; Marshak, D.R.  
J. Biol. Chem. 261, 3396-3401, 1986  
A;Title: Structural and immunological characterization of the amino-terminal domain of m  
A;Reference number: A44290; MUID:86140120; PMID:3512556  
A;Accession: A44290  
A;Molecule type: protein  
A;Residues: 20-36 <ROU>  
C;Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol  
C;Comment: Several forms of NCAM are produced by alternative splicing. See also PIR:IUMS  
C;Genetics:  
A;Gene: NCAM  
A;Map position: 9  
A;Introns: 643/3; 701/1; 770/2; 809/2; 1076/2  
C;Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu  
C;Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding; sh  
F;1-19/Domain: signal sequence #status predicted <SIG>  
F;20-1115/Product: neural cell adhesion molecule, long domain splice form #status experi  
F;20-809,1077-1115/Product: neural cell adhesion molecule, short domain splice form #sta  
F;20-711/Domain: extracellular #status predicted <EXT>  
F;34-98/Domain: immunoglobulin homology <IMM1>  
F;132-191/Domain: immunoglobulin homology <IMM2>  
F;152-156/Region: heparin binding #status predicted  
F;161-165/Region: heparin binding #status predicted  
F;228-290/Domain: immunoglobulin homology <IMM3>  
F;262-272/Region: NCAM binding #status predicted  
F;323-388/Domain: immunoglobulin homology <IMM4>  
F;420-482/Domain: immunoglobulin homology <IMM5>  
F;519-596/Domain: fibronectin type III repeat homology <FN3A>  
F;625-685/Domain: fibronectin type III repeat homology <FN3B>  
F;712-729/Domain: transmembrane #status predicted <TMM>  
F;730-1115/Domain: intracellular #status predicted <INT>  
F;41-96,139-189,235-288,330-386,427-480/Disulfide bonds: #status predicted  
F;222,316,348,424,450,479/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 3.8%; Score 132.5; DB 1; Length 1115;  
Best Local Similarity 22.5%; Pred. No. 0.73;  
Matches 83; Conservative 40; Mismatches 127; Indels 119; Gaps 17;

QY 22 LWCAAGVNLGWRPNEDDAGNGVKPAPGRD-----PL-----TCDEGDGE 62  
Db 742 LMCIA-VNLG-----KAGPGAK---GKDMEEGKAASFSDKEFIVEVTRTEERTPNH 791  
QY 63 PKSAHTSPEKKKAKELPEM--DATSSRVWLTS--TLTTSKVVIIDA-----NQP----- 108  
Db 792 DGGKHTEPNETTLPTELPADTTATVEDMLPSVTTVTTNSDITITETFATAQNSPTSETT 851  
QY 109 -----GTVVDQPTV-----CNAHVLCISSIPAASDSYPPGEMFLDSVNPEDP 152  
Db 852 TLTSSIAPPATTVPDSNSVPAGQATPSKGVTAASSSSPASAPKVAPLVDL---SDTPTSAP 908  
QY 153 GAD-----GVLAGITLVGCATRCNVPRNSCSRGDTPVLDKGQGEVATIANGKV 201  
Db 909 SASNLSSTVLANQGAVLSPSTPASAGETSKAPPASKASPAPTPTPAGAASPLAAVAAPAT 968  
QY 202 NPSQSTEEATEATEVDDPGPSE-----PETATLRPGPLTEHVFTDP----- 242  
Db 969 DAPQAKQEA-PSTKGDPEPTQPGTVKNPPEAATAPASPKSKAATTNPSQGEDLKMDEGN 1027  
QY 243 -----APTPSSGPGQ-PGSENGPEPDSSSTRPEP-----EPGSDPT 276  
Db 1028 FKTPDIDLAKDVFAALGSPRPATGASGQASELAPSPADSAPVPPAPAKTEKGPVETKSEPP 1087  
QY 277 GAGSAAAPT 285  
Db 1088 ESEAKPAPT 1096

RESULT 13  
TI8535  
high molecular mass nuclear antigen - chicken (fragment)

C;Species: Gallus gallus (chicken)  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: T18535  
R;Shimada, K.; Harata, M.; Mizuno, S.  
J. Cell Sci. 110, 3031-3041, 1997  
A;Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chicken  
A;Reference number: Z18955; MUID:9803440; PMID:9365273  
A;Accession: T18535  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-1151 <SHI>  
A;Cross-references: EMBL:D88440; NID:d1177138; PID:d1025045; PIDN:BAA24137.1

Query Match 3.8%; Score 132; DB 2; Length 1151;  
Best Local Similarity 22.5%; Pred. No. 0.83;  
Matches 73; Conservative 29; Mismatches 129; Indels 94; Gaps 12;

QY 45 KPAPGRDPLTCDREGDGEPKSAHTSPEKKKAKELPEMDATSSRVWILTSTLT----- 96  
Db 247 KPAPVTSPTIPCSSAEAKPLTAASPTASKATABAKPVPATAS---LMATKVTABAKPAPS 303  
QY 97 -----TSKVVIIDANQPGTVVDQFTVCNAHVLCISS-----IPAAASDSYPPGE 140  
Db 304 PSVPKATTDTKAVTATAPKAGPD--VKPAVAVCAEAKPAPPPPPQQLPKAAAAAPTGT 360  
QY 141 MFLDSDVNPE-DPGADGVLAGIT--LVGCATRCNVPRNSC---SSRGDTPVLDKGQ---- 190  
Db 361 ELKPATAPPHGSPRANSHVTVTVPNPVPRAAATVPTAGAVPKASTGTTTAAAPQPVPK 420  
QY 191 -----GEVATIANGKVNPSQSTEEA---TEATEVDPDGPSEPETATLRP--- 231  
Db 421 AAPVTPSPQQAQVPRATAAAAPVTPQPVTKAATTNATPPPIPKAATTTTATPVT 480  
QY 232 -----GPLETHVFTD---PAPTPSSGPGQSGSENGPEPD 261  
Db 481 QQPIPKAGTDAAPPAPVAPKAPSDGRAATGVVPAATDPQKPPPTQSVPSAVTEPKQP- 539  
QY 262 SSSTRPEPEPSGDPGTGAGSSAAPTM 286  
Db 540 ----RAAPPPSNEATPAVPSPSNLT 560

RESULT 14  
JC7303

pectate lyase (EC 4.2.2.2) - Bacillus sp.

N;Alternate names: pectate transesterase

C;Species: Bacillus sp.

C;Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 02-Sep-2000

C;Accession: JC7303

R;Ogawa, A.; Sawada, K.; Saito, K.; Hakamada, Y.; Sumitomo, N.; Hatada, Y.; Kobayashi, T  
Biosci. Biotechnol. Biochem. 64, 1133-1141, 2000

A;Title: A new high-alkaline and high-molecular-weight pectate lyase =66rom a Bacillus i

A;Reference number: JC7303

A;Accession: JC7303

A;Molecule type: DNA

A;Residues: 1-677 <OGA>

A;Cross-references: DDBJ:AB028878

A;Experimental source: strain KSM-P15

C;Genetics:

A;Gene: pel-15H

C;Keywords: calcium binding; carbon-oxygen lyase

Query Match 3.8%; Score 131; DB 2; Length 677;  
Best Local Similarity 18.3%; Pred. No. 0.48;  
Matches 108; Conservative 86; Mismatches 214; Indels 182; Gaps 26;

QY 45 KPAPGRDPLTCDREGDGEPKSAHTSPEKKKAKELPEMDATSSRVWILTSTLTTSKVVIID 104  
Db 95 QPLTGMVTVQFDWYHEGNARASSLRPMK-----IFSSASESSTTIVEIQ 139  
QY 105 ANQPGTVVDQFTVCNAHVLCISSIPAASDSYPPGEMFLDSVNPEDPGADGVLGITLV 164  
Db 140 TRDGGRHQAQ-TVNGTHLLV-----TDFAPDTWY----- 168

QY 165 GCATRCNVPRNSCSRGDTPVLVDKQGEVATIANGKVNPSQSTEEATEATEVDPGPSEP 224  
Db 169 -----RFNIVMTDTKKVD-----TYVNGELKLEQANFASTSAAQVRLKIYSQ 212  
QY 225 ETATLRPGPLTEHVF-----TDPAPTSSG-PQPGE--NGPEPDSSSTRP----- 267  
Db 213 NSPTI--GOYIDNLVYSGSNPGTDPGPGSPGTDPGTDPGTDPGSPDGTGPNP 270  
QY 268 -----BPEPSGDPTGAGSSAAPTMLGAQNGWLYVHSAVANWKCLHSIKLKDSVLSLVHV 323  
Db 271 GHPQDPGPAPEAEGDLIVAP-----NGQEG-----NPGTLNQPTTLTS-----AITRI 314  
QY 324 KGRVLVALADGTLA-----IFHRGEDG-----QWDLNHYHLM 355  
Db 315 QPGRTIYMRGGTYAFSETVLIERNNGNLEGARKRIVGYNGEKPVLDFSAQAFDPMNRGLQ 374  
QY 356 DLGHPHHSIRCMAV-----YDRVWCGYKN-KVHVIQPKTMQIEKSFDAHPRR 402  
Db 375 INGHYWH-VQIEVKEAGDNGIFIGGYNRI-----ENVETHNKDGTGLQISRYSSSATRD 429  
QY 403 ESQVRQLAWIGDVWVVSIRLDSTLRLYHAHTHQLQDVIDIEPVYSKML-GTGKLGFSFVR 461  
Db 430 E-----WPS--YNEIINVYSHNNYDPPDDGEDADGFAAKLTSGPGNV-FDGC 473  
QY 462 ITALLVAGSRLWGTGNGVVISIPLTETVTLHRGQLLGLRANKTSPTSSEGARPGGIIHV 521  
Db 474 AAYNVDDGWLTKSDTGAIYPIIRNSIAYNNGSTEG--GHSTNSDNGNGFKLGG---- 527  
QY 522 YGDDSSDRAASSFIPYCSMAQAQLCFHGHRAVKKFFVSVPGNVLATLNGS 571  
Db 528 -----SNIPVNHIVENNMAF-GNKKHGFTYNSNPGSITMTNNTS 565

RESULT 15  
T39233

probable Inositol polyphosphate phosphatase - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T39233

R:Churcher, C.M.; Gentles, S.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, August 1997

A:Reference number: Z21837  
A:Accession: T39233

A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA

A:Residues: 1-1183 <CHU>  
A:Cross-references: EMBL:Z98763; PIDN:CAB11494.1; GSPDB:GN00066; SPDB:SPAC9G1.10c

A:Experimental source: strain 972h-; cosmid c9G1  
C:Genetics:  
A:Gene: SPDB:SPAC9G1.10c

A:Map position: 1

Query Match 3.7%; Score 129; DB 2; Length 1183;  
Best Local Similarity 19.6%; Pred. No. 1.4;  
Matches 118; Conservative 81; Mismatches 272; Indels 130; Gaps 25;

QY 7 PVYCRPLVEKDPTMKLWCAAGVNLGWRPNEDDAGNGVKPAPGRDPLTCDREGDGEPKSA 66  
Db 138 PLLKRP-QQKGPEISFQSSVQSTKGNDLMKHDDTNHHQIPPP--KPNFSSKAGSSSPIS- 193  
QY 67 HTSPEKKKAKELPEMDATSSRVWILTSTLTTSKVVIIDANQPG-TVVDQFTVCNAHVLCI 125  
Db 194 -VSPLKNVKAYISQSPTHSEASSVLSSEEEENVINSSKSVPSFDLHDFFSQTFGKECPI 252  
QY 126 SSIPAA-----SDSDYPPGEMFLDSVNPEDPGADGVLAGITLVGCATRCNV----- 173  
Db 253 STAPPVLNIGDRSLETPPP---IPSPRPQPVAVEAIQOSRAVISQQLPLHVSPRKPKP 309  
QY 174 --RSNCSSRFGDTPVLVDKQGEVATIANG-KVNPSQSTEEATEATEVPD-----PGPSEPE 225  
Db 310 PLRKVSTQRSSSPIENLATKSDASLVTLGQSSPYTHIAPASEMSLIPEKPRLPSPSHTL 369

QY 226 TATLRPGPLTEHVFTDPAPT-PSSGPPQPGSENGPEPDSSTRPEPEPSGDPTGAGSSAAP 284  
Db 370 SELSSPALTSENLSKPSPLFPFPPPPRVKSLATNKPVS---MPVSTEQSDPSVAASSSSS 426  
QY 285 TMWLGAQNGWLYVHSAVANWKKC-----LHSIKLKDSVLSLVHVKGKRVLVALADGTLAIFH 340  
Db 427 SQLDVVLKGSIPDTSSVRRNPFCFVNGVESINV-DFEARIFDVSGDRLVLAGNGGLRVY- 484  
QY 341 RGEDGQWDLNHYHLMDLGHPHHSIRCMAVVYDRVWCGYKXKXVHVIQPKTMQIEKSFDAHP 400  
Db 485 ---DVTVTGLCHWH-MPLG-----DTKVTSLSPKSSP 511  
QY 401 RRESQVRQLAWIG--DG-VWVSIRLDSTLRLYHAHTHQLQDVIDIE---PYVSKMLGTGK 454  
Db 512 ENYSDDGRFVWFETRDMGLW-----EVDVQNHIVTKKSVSNCFITYVMVYKNEMWTLLD 566  
QY 455 LGFSFV-----RITALLVAGSR-LWVGTGNGVVISIPLTETVV 491  
Db 567 MGKLYVWQDEIMGLSIQSTPHSIRTIPTHATHAMVLDNRLWVWVGKSIYVYDPSTS--- 623  
QY 492 LHRGQLLGLRANKTSPTSSEGARPGGIIHVYGDSSDRAASSFIPYCSMAQAQLCFHGH 551  
Db 624 -----ENESASVLAKPMTPPGLI---GDISCGTTISNF-----TDLVIFYGHV 662  
QY 552 D 552  
Db 663 D 663

Search completed: August 23, 2004, 11:15:09  
Job time : 49 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 11:04:17 ; Search time 25 Seconds  
(without alignments)  
1355.907 Million cell updates/sec

Title: US-10-019-495-9  
Perfect score: 3473  
Sequence: 1 MKNVPVPVCRPLVEKDPTM.....VLSKAERSHIIVQVSYPTE 651

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID           | Description        |
|------------|--------|-------------|--------|--------------|--------------------|
| 1          | 3473   | 100.0       | 1334   | 1 JIP3 HUMAN | Q9upt6 homo sapien |
| 2          | 3191   | 91.9        | 1337   | 1 JIP3 MOUSE | Q9esn9 m c-jun-ami |
| 3          | 1217.5 | 35.1        | 1227   | 1 JIP3 DROME | Q9ggf1 drosophila  |
| 4          | 1152.5 | 33.2        | 1157   | 1 JIP CAEL   | P34609 caenorhabdi |
| 5          | 183.5  | 5.3         | 1222   | 1 YNG2 CAEL  | Q21653 caenorhabdi |
| 6          | 161    | 4.6         | 797    | 1 SHK3 HUMAN | Q9byb0 homo sapien |
| 7          | 146.5  | 4.2         | 1815   | 1 SHK3 RAT   | Q9jju4 rattus norv |
| 8          | 138    | 4.0         | 1367   | 1 AMYH YEAST | P08640 saccharomyc |
| 9          | 137.5  | 4.0         | 634    | 1 HWP1 CANAL | P46593 candida alb |
| 10         | 136    | 3.9         | 3644   | 1 MINT MOUSE | Q62504 mus musculu |
| 11         | 132.5  | 3.8         | 1115   | 1 NCAL MOUSE | P13595 mus musculu |
| 12         | 130    | 3.7         | 677    | 1 OGFR HUMAN | Q9nzt2 homo sapien |
| 13         | 129    | 3.7         | 1664   | 1 SLP1 CLOTH | Q06852 clostridium |
| 14         | 128    | 3.7         | 1567   | 1 FMN2 MOUSE | Q9jl04 mus musculu |
| 15         | 126.5  | 3.6         | 742    | 1 PKWA THECU | P49695 thermomonos |
| 16         | 125.5  | 3.6         | 442    | 1 SDC3 MOUSE | Q64519 mus musculu |
| 17         | 124.5  | 3.6         | 535    | 1 SPKC SYN3  | P74745 synecocyst  |
| 18         | 123.5  | 3.6         | 824    | 1 JIP2 HUMAN | Q13387 homo sapien |
| 19         | 123.5  | 3.6         | 886    | 1 SM6B MOUSE | Q54951 mus musculu |
| 20         | 122.5  | 3.5         | 279    | 1 Y091 NPVOP | O10341 orgyia pseu |
| 21         | 122    | 3.5         | 1225   | 1 DAT1 HUMAN | Q9btc0 homo sapien |
| 22         | 121.5  | 3.5         | 143    | 1 PAR1 TRYBB | P08469 trypanosoma |
| 23         | 121.5  | 3.5         | 145    | 1 PARC TRYBB | Q06084 trypanosoma |
| 24         | 121.5  | 3.5         | 2441   | 1 CBP MOUSE  | P45481 mus musculu |
| 25         | 121.5  | 3.5         | 5147   | 1 PCLO HUMAN | Q9y6v0 homo sapien |
| 26         | 120.5  | 3.5         | 565    | 1 MOT8 MOUSE | O70324 mus musculu |
| 27         | 120.5  | 3.5         | 1134   | 1 ANS1 HUMAN | Q92625 homo sapien |
| 28         | 119.5  | 3.4         | 667    | 1 SIX5 MOUSE | P70178 mus musculu |
| 29         | 119.5  | 3.4         | 1520   | 1 ACFD ECOLI | Q46837 escherichia |
| 30         | 119    | 3.4         | 720    | 1 Z408 HUMAN | Q9h9d4 homo sapien |
| 31         | 119    | 3.4         | 2415   | 1 PGCA HUMAN | P16112 homo sapien |
| 32         | 118.5  | 3.4         | 797    | 1 VGLX HSVEB | P28968 equine herp |
| 33         | 118.5  | 3.4         | 1101   | 1 GUNC CELFI | P14090 cellulomona |

|            |   |           |               |
|------------|---|-----------|---------------|
| RESULT 1   |   |           |               |
| JIP3_HUMAN |   |           |               |
| ID         | JIP3_HUMAN  | STANDARD; | PRT; 1334 AA. |
| AC         | Q9UPT6; Q96RY4; Q9H4I4; Q9H7P1; Q9NUG0;   |           |               |
| DT         | 28-FEB-2003 (Rel. 41, Created)  |           |               |
| DT         | 28-FEB-2003 (Rel. 41, Last sequence update)   |           |               |
| DT         | 10-OCT-2003 (Rel. 42, Last annotation update)   |           |               |
| DE         | C-jun-amino-terminal kinase interacting protein 3 (JNK-interacting protein 3) (JIP-3) (JNK MAP kinase scaffold protein 3) (Mitogen-activated protein kinase 8-interacting protein 3).   |           |               |
| GN         | MAPK8IP3 OR JIP3 OR KIAA1066.   |           |               |
| OS         | Homo sapiens (Human).   |           |               |
| OC         | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;   |           |               |
| OC         | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  |           |               |
| OX         | NCBI_TaxID=9606;  |           |               |
| RN         | [1]   |           |               |
| RP         | SEQUENCE FROM N.A.  |           |               |
| RX         | MEDLINE=21096910; PubMed=11157797;  |           |               |
| RA         | Daniels R.J., Peden J.F., Lloyd C., Horsley S.W., Clark K., Tufarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J., Higgs D.R.;  |           |               |
| RA         | "Sequence, structure and pathology of the fully annotated terminal 2 Mb of the short arm of human chromosome 16.";  |           |               |
| RL         | Hum. Mol. Genet. 10:339-352(2001).  |           |               |
| RN         | [2]   |           |               |
| RP         | SEQUENCE FROM N.A.  |           |               |
| RA         | Bagguley C., Hall R.;   |           |               |
| RL         | Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.   |           |               |
| RN         | [3]   |           |               |
| RP         | SEQUENCE FROM N.A., AND REVISIONS TO 322; 412 AND 751.  |           |               |
| RC         | TISSUE=Brain;   |           |               |
| RA         | Ohara O., Nagase T., Kikuno R.;   |           |               |
| RL         | Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.   |           |               |
| RN         | [4]   |           |               |
| RP         | SEQUENCE OF 266-1334 FROM N.A.  |           |               |
| RC         | TISSUE=Brain;   |           |               |
| RX         | MEDLINE=99397452; PubMed=10470851;  |           |               |
| RA         | Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  |           |               |
| RT         | "Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";  |           |               |
| RL         | DNA Res. 6:197-205(1999).   |           |               |
| RN         | [5]   |           |               |
| RP         | SEQUENCE OF 1-252 FROM N.A.   |           |               |
| RC         | TISSUE=Spleen;  |           |               |
| RA         | Ohara O., Nagase T., Kikuno R., Okumura K.;   |           |               |
| RT         | "The nucleotide sequence of a long cDNA clone isolated from human spleen.";   |           |               |
| RL         | Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.   |           |               |
| CC         | !- FUNCTION: The JNK-interacting protein (JIP) group of scaffold proteins selectively mediates JNK signaling by aggregating specific components of the MAPK cascade to form a functional JNK signaling module. May function as a regulator of vesicle transport, through interactions with the JNK-signaling components and motor proteins (By similarity). |           |               |

Q9w0k7 drosophila  
O88573 mus musculu  
P22534 caldocellum  
Q28181 bos taurus  
Q10383 mycobacteri  
Q9y2n3 homo sapien  
P53564 mus musculu  
Q9eqn3 mus musculu  
P97260 cricetus  
Q9fpq6 chlamydomon  
P40200 homo sapien  
P41245 mus musculu

ALIGNMENTS



RA Kelkar N., Gupta S., Dickens M., Davis R.J.;  
RT "Interaction of a mitogen-activated protein kinase signaling module  
RL with the neuronal protein JIP3.";  
RL Mol. Cell. Biol. 20:1030-1043(2000).  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORM 1C), SUBCELLULAR LOCATION, AND INTERACTION  
RP WITH KLC1.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RX MEDLINE=20560743; PubMed=11106729;  
RA Bowman A.B., Kamal A., Ritchings B.W., Philp A.V., McGrail M.,  
RA Gindhart J.G., Goldstein L.S.B.;  
RT "Kinesin-dependent axonal transport is mediated by the Sunday Driver  
RT (SYD) protein.";  
RL Cell 103:583-594(2000).  
RN [5]  
RP SEQUENCE OF 1240-1337 FROM N.A.  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer M.F., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo A.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [6]  
RP INTERACTION WITH KLC.  
RC TISSUE=Brain;  
RX MEDLINE=21135887; PubMed=11238452;  
RA Verhey K.J., Meyer D., Deehan R., Blenis J., Schnapp B.J.,  
RA Rapoport T.A., Margolis B.;  
RT "Cargo of kinesin identified as JIP scaffolding proteins and  
RT associated signaling molecules.";  
RL J. Cell Biol. 152:959-970(2001).  
CC -!- FUNCTION: The JNK-interacting protein (JIP) group of scaffold  
CC proteins selectively mediates JNK signaling by aggregating  
CC specific components of the MAPK cascade to form a functional JNK  
CC signaling module. May function as a regulator of vesicle  
CC transport, through interactions with the JNK-signaling components  
CC and motor proteins.  
CC -!- SUBUNIT: Forms homo- or heterooligomeric complexes. The central  
CC region of Mapk8ip3 interacts with the C-terminal of Mapk8ip2 but  
CC not Mapk8ip1. Binds specific components of the JNK signaling  
CC pathway namely Mapk8, Mapk9 and Mapk10 to the N-terminal region,  
CC Map2k4 and Map2k7 to the central region and Map3k11 to the C-  
CC terminal region. Binds the TPR motif-containing C-terminal of  
CC kinesin light chain, pre-assembled Mapk8ip1 scaffolding complexes  
CC are then transported as a cargo of kinesin, to the required  
CC subcellular location.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; localised in the soma and  
CC growth cones of differentiated neurites and the Golgi and vesicles  
CC of the early secretory compartment of epithelial cells.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=6;  
CC Name=lc; Synonyms=3b;  
CC IsoId=Q9ESN9-1; Sequence=Displayed;  
CC Name=1a;  
CC IsoId=Q9ESN9-2; Sequence=VSP\_002775, VSP\_002777;  
CC Name=1b;  
CC IsoId=Q9ESN9-3; Sequence=VSP\_002776, VSP\_002777;  
CC Name=1d;  
CC

CC IsoId=Q9ESN9-4; Sequence=VSP\_002775;  
CC Name=3a;  
CC IsoId=Q9ESN9-5; Sequence=VSP\_002778, VSP\_002779;  
CC Name=1e;  
CC IsoId=Q9ESN9-6; Sequence=VSP\_002776;  
CC TISSUE SPECIFICITY: Highly expressed throughout many regions of  
CC the brain and at lower levels in the heart, liver, lung, testes  
CC and kidney. All isoforms have been identified in the brain,  
CC Mapk8ip3A is also expressed in the spleen and lung.  
CC -!- INDUCTION: Expressed in neurites 5 days following initiation of  
CC nerve growth factor Ngf induced differentiation. Ngf withdrawal  
CC results in the down-regulation of Mapk8ip3 protein by caspase-  
CC mediated cleavage.  
CC -!- SIMILARITY: Belongs to the JIP scaffold family.  
CC [5]  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AB005662; BAA85874.1; -  
CC EMBL; AB043124; BAB16675.1; -  
CC EMBL; AB043125; BAB16676.1; -  
CC EMBL; AB043123; BAB16674.1; -  
CC EMBL; AB043129; BAB16685.1; -  
CC EMBL; AB043126; BAB16685.1; JOINED.  
CC EMBL; AB043127; BAB16685.1; JOINED.  
CC EMBL; AB043128; BAB16685.1; JOINED.  
CC EMBL; AF178637; AAF26843.1; -  
CC EMBL; AF178636; AAF26842.1; -  
CC EMBL; AF262046; AAG36931.1; -  
CC EMBL; BC004003; AAH04003.1; -  
CC MGD; MGI:1353598; Mapk8ip3.  
CC GO; GO:0005737; C:cytoplasm; IEP.  
CC GO; GO:0019894; F:kinesin binding; IPI.  
CC GO; GO:0005078; F:MAP-kinase scaffold activity; IPI.  
CC GO; GO:0019901; F:protein kinase binding; IPI.  
CC GO; GO:0046328; P:regulation of JNK cascade; IDA.  
CC GO; GO:0016192; P:vesicle-mediated transport; IDA.  
CC KW Alternative splicing; Phosphorylation; Coiled coil.  
CC DOMAIN 58 177  
CC COILED COIL (POTENTIAL).  
CC FT MOD\_RES 437 555  
CC MOD\_RES 266 266  
CC MOD\_RES 276 276  
CC MOD\_RES 287 287  
CC VARSPLIC 201 201  
CC FT VARSPLIC 201 201  
CC S -> SPQSWRKS (in isoform 1b and isoform  
CC 1e).  
CC /FTid=VSP\_002776.  
CC Missing (in isoform 1a and isoform 1b).  
CC /FTid=VSP\_002777.  
CC Missing (in isoform 3a).  
CC /FTid=VSP\_002778.  
CC Missing (in isoform 3a).  
CC /FTid=VSP\_002779.  
CC R->G: RESULTS IN INHIBITION OF JNK  
CC BINDING.  
CC P->G: RESULTS IN INHIBITION OF JNK  
CC BINDING.  
CC T->G: RESULTS IN INHIBITION OF JNK  
CC BINDING.  
CC S->G: RESULTS IN INHIBITION OF JNK  
CC BINDING.  
CC L->G: RESULTS IN INHIBITION OF JNK  
CC BINDING.  
CC T->A: RESULTS IN LOSS OF PHOSPHORYLATION  
CC OF MAPK8IP3; WHEN ASSOCIATED WITH A-276  
CC AND 287. DOES NOT EFFECT BINDING OF  
CC COMPONENTS OF THE JNK PATHWAY.

|                       |   |                                 |                          |   |                   |
|-----------------------|---|---------------------------------|--------------------------|---|-------------------|
| FT                    | MUTAGEN   | 276                             | 276                      | T->A: RESULTS IN LOSS OF PHOSPHORYLATION OF MAPK8IP3; WHEN ASSOCIATED WITH A-266 AND A-287. DOES NOT EFFECT BINDING OF COMPONENTS OF THE JNK PATHWAY. |                   |
| FT                    |   |                                 |                          | T->A: RESULTS IN LOSS OF PHOSPHORYLATION OF MAPK8IP3; WHEN ASSOCIATED WITH A-266 AND A-287. DOES NOT EFFECT BINDING OF COMPONENTS OF THE JNK PATHWAY. |                   |
| FT                    | CONFLICT  | 312                             | 312                      | K -> R (IN REF. 4).   |                   |
| FT                    | CONFLICT  | 376                             | 376                      | F -> L (IN REF. 3; AAF26843).   |                   |
| FT                    | CONFLICT  | 561                             | 561                      | E -> K (IN REF. 4).   |                   |
| Query Match           |   |                                 |                          |   |                   |
| Best Local Similarity |   | 91.9%;                          | Score 3191;              | DB 1;   | Length 1337;      |
| Matches 594;          |   | Conservative                    | 24;                      | Mismatches 33;  | Indels 0; Gaps 0; |
| Qy                    | 1   | MKNVPVPVYCRPLVEKDP              | TMKLWCAAGVNL             | SGWRPNEDDAGNGVKPAPGRDPLTCDREGD  | 60                |
| Db                    | 687   | MKNVPVPVYCRPLVEKDP              | STKLWCAAGVNL             | SGWKPHCEEDSSNGPKVPVGRDPLTCDREGE   | 746               |
| Qy                    | 61  | GEPKSAHTSPEKKAKELPEMDATSSRVWIL  | STLTTSKVVIIDANQP         | GTVVDDQFTVCNA   | 120               |
| Db                    | 747   | GEPKSTHPSPEKKAKETPEADATSSRVWIL  | STLTTSKVVIIDANQP         | GTIVDQFTVCNA  | 806               |
| Qy                    | 121   | HVLCISSIPAASDSYPPGEMFLDS        | VNPEDPGADGVL             | AGITLVGCATRCNVPRNSCSSR  | 180               |
| Db                    | 807   | HVLCISSIPAASDSYPPGEMFLDS        | VNPEDSGADGVL             | AGITLVGCATRCNVPRNSCSSR  | 866               |
| Qy                    | 181   | GDPVLDKGQGEVATIANGKVNPSQSTEEATE | VPDGPSEPETATLRPGPLTEHVFT |   | 240               |
| Db                    | 867   | GDPVLDKGQGVATTANGKVNPSQSTEEATE  | VPDGPSESEATTVRPGPLTEHVFT |   | 926               |
| Qy                    | 241   | DPAPTSSGPPQGSSENGPEPDSSTRPEPE   | PSGDPGTGAGSSAAPT         | MWLGAQNGWLYVHSA   | 300               |
| Db                    | 927   | DPAPTSSSTQPASENGSENGTIVQPQVEP   | SGELSTTTSSAAPT           | MWLGAQNGWLYVHSA   | 986               |
| Qy                    | 301   | VANWKKCLHSIKLKDSVLSLVHVKGRVL    | VALADGTLAIFHRGEDG        | QWDL  | SNYHLMDLGHP       |
| Db                    | 987   | VANWKKCLHSIKLKDSVLSLVHVKGRVL    | VALADGTLAIFHRGEDG        | QWDL  | SNYHLMDLGHP       |
| Qy                    | 361   | HHSIRCMVAVYDRVWCGYKKNKHVVIQ     | PKTMQIEKSFDAHP           | PRRESQVRQLAWIGDVVWSI  | 420               |
| Db                    | 1047  | HHSIRCMVAVNDRVWCGYKKNKHVVIQ     | PKTMQIEKSFDAHP           | PRRESQVRQLAWIGDVVWSI  | 1106              |
| Qy                    | 421   | RLDSTLRLYHAHTHOHLQDV            | DIIEPVYSKMLGTGKLGFSFVRIT | ALLVAGSRLWVG  | TGNGV             |
| Db                    | 1107  | RLDSTLRLYHAHTHOHLQDV            | DIIEPVYSKMLGTGKLGFSFVRIT | ALLIAGNRLWVG  | TGNGV             |
| Qy                    | 481   | VISIPLTETVTLHRLGQLLGLRANKTSPT   | SGEGARPGGIIHVYGD         | SSDRAASSFIPYCSM   | 540               |
| Db                    | 1167  | VISIPLTETVTLHRLGQLLGLRANKTSPT   | SGEGTRPGGIIHVYGD         | SSDKAASSFIPYCSM   | 1226              |
| Qy                    | 541   | AQAQLCFHGHHRDAVKFFVSVPGNV       | LATLNGSVLDS              | PAEGPGPAAPASEVEGQKLRNVL   | VL                |
| Db                    | 1227  | AQAQLCFHGHHRDAVKFFVSVPGNV       | LATLNGSVLDS              | PEGPGPAAPADAEGQKLRNVL   | VL                |
| Qy                    | 601   | SGGEGYIDFRIGDGEDDETEEGAGDMSQ    | VKPVLSKAERSHII           | VWQVSYTPE   | 651               |
| Db                    | 1287  | SGGEGYIDFRIGDGEDDETEECAGDV      | NQTKPSLSKAERSHII         | VWQVSYTPE   | 1337              |
| RESULT 3              |   |                                 |                          |   |                   |
| ID                    | JIP3_DROME  | STANDARD;                       | PRT;                     | 1227  | AA.               |
| AC                    | Q9GQF1; Q95SK0; Q9VSC0;                                       |                                 |                          |   |                   |
| DT                    | 28-FEB-2003 (Rel. 41, Created)                                |                                 |                          |   |                   |
| DT                    | 28-FEB-2003 (Rel. 41, Last sequence update)                   |                                 |                          |   |                   |
| DT                    | 10-OCT-2003 (Rel. 42, Last annotation update)                 |                                 |                          |   |                   |
| DE                    | JNK-interacting protein 3 (Sunday driver protein).            |                                 |                          |   |                   |
| GN                    | SYD OR CG8110.  |                                 |                          |   |                   |
| OS                    | Drosophila melanogaster (Fruit fly).                          |                                 |                          |   |                   |
| OC                    | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; |                                 |                          |   |                   |
| OC                    | Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;    |                                 |                          |   |                   |

|    |  |
|----|--|
| OC | Ephydroidea; Drosophilidae; Drosophila.  |
| OX | NCBI_TaxID=7227;   |
| RP | [1]  |
| RX | SEQUENCE FROM N.A. (ISOFORM LONG), AND INTERACTION WITH KLC.   |
| RA | MEDLINE=20560743; PubMed=11106729;   |
| RA | Bowman A.R., Kamai A., Ritchings B.W., Philp A.V., McGrail M., Gindhart J.G., Goldstein L.S.B.;  |
| RT | "Kinesin-dependent axonal transport is mediated by the sunday driver (SYD) protein.";  |
| RL | Cell 103:583-594 (2000).   |
| RP | [2]  |
| RX | SEQUENCE FROM N.A.   |
| RC | STRAIN=Berkeley;   |
| RX | MEDLINE=20196006; PubMed=10731132;   |
| RA | Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Adams M.D., Celniker S.E., Li P.W., Hoskins R.A., Galle R.F., Amanatides P.G., Scherer S.E., Richards S., Ashburner M., Henderson S.N., George R.A., Lewis S.E., Yandell M.D., Zhang Q., Chen L.X., Sutton G.G., Wortman J.R., Yandell M.D., Zhang M., Pfeiffer B.D., Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Miklos G.L.G., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Baldwin D., Abril J.F., Agbayani A., An H.-J., Andrews-Pfammkoch C., Beasley E.M., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Berman B.P., Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W., Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; |
| RT | "The genome sequence of Drosophila melanogaster.";   |
| RL | Science 287:2185-2195 (2000).  |
| RP | [3]  |
| RX | REVISIONS, AND ALTERNATIVE SPLICING.   |
| RX | MEDLINE=22426069; PubMed=12537572;   |
| RA | Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S., Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E., Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P., Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A., Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q., Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M., Lewis S.E.;   |
| RT | "Annotation of the Drosophila melanogaster euchromatic genome: a systematic review.";  |
| RL | Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).  |
| RP | [4]  |
| RX | SEQUENCE OF 937-1227 FROM N.A.   |
| RC | STRAIN=Berkeley; TISSUE=Head;  |
| RX | MEDLINE=22426066; PubMed=12537569;   |
| RA | Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M., George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H., Rubin G.M., Celniker S.E.;  |
| RT | "A Drosophila full-length cDNA resource.";   |

Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
-!- FUNCTION: The JNK-interacting protein (JIP) group of scaffold proteins selectively mediates JNK-signaling by aggregating specific components of the MAPK cascade to form a functional JNK signaling module. May function as a regulator of vesicle transport, through interactions with the JNK-signaling components and motor proteins. Syd is required for efficient kinesin-I mediated axonal transport.  
-!- SUBUNIT: Forms homo- and heterooligomeric complexes. Binds the tpr motif-containing C-terminal of kinesin light chain, Klc. Pre-assembled syd scaffolding complexes are then transported as a cargo of kinesin, to the required subcellular location.  
-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
-!- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=2;  
Name=Long;  
IsoId=Q9GQF1-1; Sequence=Displayed;  
Name=Short;  
IsoId=Q9GQF1-2; Sequence=VSP\_002780;  
Note=No experimental confirmation available;  
-!- SIMILARITY: Belongs to the JIP scaffold family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AF262045; AAC36930.1; -  
EMBL; AE003557; AAF50505.2; -  
EMBL; AE003557; AAN12032.1; -  
EMBL; AY060748; AAL28296.1; -  
FlyBase; FBgn0024187; syd.  
GO; GO:0005737; Cytoplasm; ISS.  
GO; GO:0030140; C.trans-Golgi network transport vesicle; IDA.  
GO; GO:0019894; F.kinesin binding; IPI.  
GO; GO:0005078; F.MAP-kinase scaffold activity; ISS.  
GO; GO:0019901; F.protein kinase binding; ISS.  
GO; GO:0008088; P.axon cargo transport; IMP.  
GO; GO:0046328; P.regulation of JNK cascade; ISS.  
GO; GO:0016192; P.vesicle-mediated transport; IEP.  
Coiled coil; Alternative splicing.  
KW DOMAIN 84 191 COILED COIL (POTENTIAL).  
DOMAIN 363 489 COILED COIL (POTENTIAL).  
DOMAIN 814 849 COILED COIL (POTENTIAL).  
VARSP LIC 1 55 Missing (in isoform Short).  
FT CONFLICT 983 983 S -> G (IN REF. 3).  
FT SEQUENCE 1227 AA; 136711 MW; 2162C18066A057B8 CRC64;  
Query Match 35.1%; Score 1217.5; DB 1; Length 1227;  
Best Local Similarity 41.1%; Pred. No. 8e-67;  
Matches 287; Conservative 70; Mismatches 174; Indels 167; Gaps 20;  
4 VPVPVYCRPLVEKPTMKLWCAAGVNLGWRPNEDDAGNGVKPAPGRDPLTCDREGDGE 63  
641 VPVPVYCNPLAEASPHMKVFCAGVNLHG-----GFTKNGQSLIPANSFYA-----P 687  
64 KSA-----HTSPEKKKAKEL-----PEMDATSSRVWILTSTLTTSKVVID 104  
688 KSTLKIAEITSPADQSMERALDRQMARVSLTLEPETQ-LSSFVWICTSTHAASTVSVD 746  
105 ANQPGTVVDQFTVCNAHVLCTISSIPAASDSY-----PPGEMFLSDVNPEDPGAD 155  
747 ANQSATVLDAPFPCASHLLCIASVQGMESDYALLEQSEVVVKAGEML-----QRPG-- 797  
156 GVLAGITLVGCATRCNVPNSNCSSRGDTPVLDKGQGEVATIANGKVNPSQSTEEATEATE 215  
798 ---EGTELLGKVEFVRV-----KPKSDDEQ-----NSNEKQQQEEEEEKATE 837  
216 -----VPDPGPSEP-----ETATLR---PGPLTEHVFTDPAPTSSGPGQSGENGPEDSS 263

Db 838 KSNEQLPAVSAEPLGNVEAIKIQPLPGA-----PQRLSTDGNQNNNNSSSSN 889  
QY 264 -----STRPEPSGDPTGAGSSAAPTMLGAQNGWLYVHSAVANWKKCLHSIKLDSV 317  
Db 890 LLFATKSLNPILETKDRDEPAMSSVGPTMLGAQDGWLYVHSSVGRWHECLHRLPDAV 949  
QY 318 LSLVHVKGRVLVALADGTLAIFHRGEDGQWDLNHYHMLDLGHPHHSIRCMVAVVDRVWCG 377  
Db 950 LAIVHVEARVVVALANAQLAVFRQTDGQWDLNHYHMLDLGHPHHSIRCLCVAGERIWA 1009  
QY 378 YKNKVHVIOPTQWIEKSFDAHPRESQVRLAWIGDVVWVSIKLDSTLRLYHAHTHQL 437  
Db 1010 HRNKIFIVDPVSLNIVHSLDAHPKESQVQMAATGAGVWVSIKLDSTLRLYNTHTFEHK 1069  
QY 438 QDVDIEPYVSKMLGTGKLGFSFVRITALLVAGSRLWVGTVNGVVISIPLTETVVLHRGQL 497  
Db 1070 QDVDIEPYVSKMLGTGKLGFSFVRITALLVAGSRLWVGTVNGVVISIPLTETVVLHRGQL 497  
QY 498 LGLRANKTSPTSSEGARPGGIHVYDSDSDRAASSFIPYCSMAQAQLCFHGRDAVKFF 557  
Db 1121 --VQPKSSSDPHGQ-----MPLCCMANAQLSFHGRDAVKFF 1155  
QY 558 VSVPGNVLATLNGSVLDSPAEGPGPAAPASEVEGQKLRNVLVSGGEGYIDFRIGDG--- 614  
Db 1156 VSVPMQLQPNLNGGLTFT-----NKRPMVLVCGGEGYIDFRINDNDME 1199  
QY 615 -----EDDETEEGAGDMSQVKPVLKASRSHIIVQVS 647  
Db 1200 NSIQLEPNQTIENRGD-----KSYLIVWHVS 1225  
RESULT 4  
JIP\_CAEEL  
ID JIP\_CAEEL STANDARD; PRT; 1157 AA.  
AC P34609; Q95V72;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE JNK-interacting protein (JIP) (JNK MAP kinase scaffold protein)  
DE (Uncoordinated protein 16).  
GN UNC-16 OR ZK1098.10.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM C), FUNCTION, INTERACTION WITH JNK-1;  
RP JNK-1 AND SEK-1, SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.  
RX MEDLINE=99321749; PubMed=10393177;  
RA Kawasaki M., Hisamoto N., Iino Y., Yamamoto M., Ninomiya-Tsuji J.,  
RA Matsumoto K.;  
RT "A Caenorhabditis elegans JNK signal transduction pathway regulates  
RT coordinated movement via type-D GABAergic motor neurons.";  
RL EMBO J. 18:3604-3615(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fraser A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,  
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,  
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,  
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
RA Wohldman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).



```
CC CC      -!- SIMILARITY: Contains 1 DEL-homology (DH) domain.
CC CC      -----
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----
CC CC      EMBL; U00055; AAL06044.3; --
CC CC      WormPep; R02F2.2; CE31046.
CC CC      InterPro; IPR001331; GDS CDC24.
CC CC      InterPro; IPR000219; RhoGEF.
CC CC      Pfam; PF00621; RhoGEF; 1.
CC CC      SMART; SM00325; RhoGEF; 1.
CC CC      PROSITE; PS00741; DH_1; 1.
CC CC      PROSITE; PS001010; DH_2; 1.
CC CC      KW Hypothetical protein; Signal.
CC CC      FT SIGNAL 1 15 POTENTIAL.
CC CC      CHAIN 16 1222 HYPOTHETICAL PROTEIN R02F2.2.
CC CC      DOMAIN 83 118 ARG-RICH.
CC CC      DOMAIN 397 584 DH.
CC CC      SEQUENCE 1222 AA; 137354 MW; C12D723AF9D72973 CRC64;
CC CC
CC CC      Query Match 5.3%; Score 183.5; DB 1; Length 1222;
CC CC      Best Local Similarity 20.4%; Pred. No. 0.0011;
CC CC      Matches 107; Conservative 80; Mismatches 193; Indels 144; Gaps 23;
CC CC
QY 162 TLVGCATR-----CN---VPRSNCSRGDTPVL-----DKQGEVATIANGKVNPSQSTE 208
Db 805 TIIAHQTRPGLQLCTATFVPGKRVDS---TPSLWVCASDFSGQVAVMA---LDTGEITI 858
QY 209 EATEA-----TEVPDPGPSEPETATLRPGPLTEHVFTDPAPTSSGGPQSGENGPEPD 261
Db 859 ESCSAIGNAAVTAMCTVPPPMKLRKRIKSKSLEHL-----NETIMD 902
QY 262 SSSTRPEPEPGDPTGAGSSAAPTMLGQNGWLYVHSAVANWKKCLHS--IKLKDSVLS 319
Db 903 INSSGSDTESSSD-EGTSTAGQTTWIGNDDGEVFNSTERVSRARDRLARLNSITS 961
QY 320 LVHVKGRLVALADGT---LAIHFRGEDQWDLNLYHMLDGLPHHS-IRCMVVYDVRW 375
Db 962 ICAANGNVLVATSYSNQVQLLFRPASDGSWDLN--PQTGVHVCQAPITSMQLIGRRVI 1019
QY 376 CGYKKNKHVIQPKTMQIEKSFDAHPRRESQVRQLAWI--GDGVVVSIRLDSTLRLYHAH 434
Db 1020 IASGNWLHAYFVDVTGKFPPEILP--SSDVITLMYVTGNVFLCGRKSTEVFVDVFN 1077
QY 435 OHLQVDVIEPYVSKMLG-----TGKLGFSFVRITALLVAGSRLWVGNGVVISPLT 487
Db 1078 SIINHFNVSFVRSQSLSGREHILREHKG--CLRISCLTVARSHLWIGTSAGCVLS---- 1131
QY 488 ETVVLHRLGLLRANKTSPTSSEGEARPGGIIHVYGDSSDRAASSFIPYCSMAQAQLCF 547
Db 1132 -----TSVQSARSQPT-----PDLRVE 1149
QY 548 HGRDAVKFFVSVPGNVLATNGSVLDSPAEGGPA---APASEVEGQKLRNVLVLSGG 603
Db 1150 IGH-----SGPCRILLPVH-----TPSHSNHPSKQKRSSLNVPQAQSSQLMLVSCG 1196
QY 604 EGYIDFRIGDGEDDETEGAGDMSQVKPVLKSAERSHIVQVS 647
Db 1197 EGLDD---GTATQDPSIDAI-----NHLIFWKCS 1222
CC CC
RESULT 6
SHK3_HUMAN
ID SHK3_HUMAN STANDARD; PRT; 797 AA.
AC Q9BYB0; Q8TET3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
```

```
DE SH3 and multiple ankyrin repeat domains protein 3 (Shank3) (Proline-
DE rich synapse-associated protein 2) (ProSAP2) (Fragment).
GN SHANK3 OR PSAP2 OR KIAA1650.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21156230; PubMed=11258795;
RA Hirose M., Nagase T., Murahashi Y., Kikuno R., Ohara O.;
RT "Identification of novel transcribed sequences on human chromosome 22
RT by expressed sequence tag mapping.";
RL DNA Res. 8:1-9(2001).
RN [2]
RP SEQUENCE OF 28-797 FROM N.A.
RC TISSUE=Spleen;
RA Ohara O., Nagase T., Kikuno R., Okumura K.;
RT "The nucleotide sequence of a long cDNA clone isolated from human
RT spleen.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHROMOSOMAL TRANSLOCATION.
RX MEDLINE=21344586; PubMed=11431708;
RA Bonaglia M.C., Giorda R., Borgatti R., Felisari G., Gagliardi C.,
RA Selicorni A., Zuffardi O.;
RT "Disruption of the ProSAP2 gene in a t(12;22)(q24.1;q13.3) is
RT associated with the 22q13.3 deletion syndrome.";
RL Am. J. Hum. Genet. 69:261-268(2001).
RN [4]
RP REVIEW.
RX MEDLINE=20267867; PubMed=10806096;
RA Sheng M., Kim E.;
RT "The Shank family of scaffold proteins.";
RL J. Cell Sci. 113:1851-1856(2000).
CC -!- FUNCTION: Seems to be an adapter protein in the postsynaptic
CC density (PSD) of excitatory synapses that interconnects receptors
CC of the postsynaptic membrane including NMDA-type and metabotropic
CC glutamate receptors via complexes with GKAP/PSD-95 and Homer,
CC respectively, and the actin-based cytoskeleton. May play a role in
CC the structural and functional organization of the dendritic spine
CC and synaptic junction.
CC -!- SUBUNIT: May homomultimerize via its SAM domain (By similarity).
CC Interacts with DLGAP1/GKAP, MGLUR1A, MGLUR5 C-termini via its PDZ
CC domain (By similarity). Interacts with Homer-1, Homer-2, Homer-3
CC and CCTN/cortactin SH3 domain (By similarity). Is part of a
CC complex with DLG4/PSD-95 and DLGAP1/GKAP (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of
CC neuronal cells (By similarity).
CC -!- TISSUE SPECIFICITY: Expressed in the cerebral cortex and the
CC cerebellum.
CC -!- DISEASE: Involved in the chromosome 22q13.3 deletion syndrome by a
CC chromosomal translocation t(12;22)(q24.1;q13.3) that involves
CC DIP13B/FLJ10659 and SHANK3.
CC -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB051437; BAB33320.1; --
CC EMBL; AK074038; BAB84864.1; --
CC Genew; HGNC:14294; SHANK3.
CC MIM; 606230; --
CC MIM; 606232; --
CC InterPro; IPR001660; SAM.
CC Pfam; PF00536; SAM; 1.
CC SMART; SM00454; SAM; 1.
CC PROSITE; PS50105; SAM_DOMAIN; 1.
```

```
KW SH3-binding; Coiled coil; Chromosomal translocation.
FT NON_TER 1
FT SITE 476 482 SH3-BINDING (POTENTIAL).
FT DOMAIN 734 797 SAM.
FT DOMAIN 560 580 COILED COIL (POTENTIAL).
SQ SEQUENCE 797 AA; 82958 MW; CB94383ABC1390E6 CRC64;

Query Match 4.6%; Score 161; DB 1; Length 797;
Best Local Similarity 20.7%; Pred. No. 0.015;
Matches 138; Conservative 86; Mismatches 290; Indels 152; Gaps 28;

QY 37 EDDAGNGVKPA-----PGRDPLTCDREGDGEPSAHTSPEKKKAKELPEMDATSSR---- 87
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 216 DPERGSLASPAFSRSPAWIPVAPRAREAEKVPERKSPEDKKSMLSVLDTSLQRPAGL 275
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
QY 88 --WILTSLTTSKWIIDANQPGTVVDQFTVCNAHVLCISSIPAAASDSYPPGEMFLDS 145
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 276 IVVHATSNQGFSLRGAAEEHPGT--PELAAPMQSAVAE-PLSPRAQPPGG--TPA 330
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
QY 146 DVNP-----EDPGADGVLAGITLVGCATRCNVPRSNCSRGDTPVLDKQGEVATIANG 199
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 331 DAGPGQGSSEEP--ELVFA-----VNLPPAQLSSSD-----EETREELARI--G 371
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
QY 200 KVNPSQS--TEATEATEVPDPGPSEPETATLPGPLTEHVFTDPATPSSGPPGSENGP 258
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 372 LVPPPEEFANGVLLATPLAGPFS--PTTV---PSPASGKPPSSPPPPAPESAADSGVE--- 424
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
QY 259 EPDSSSTRPEPSPGDPGTGAGSSAAPTMLGAQNGWLY-VHSAVANWKKCLHKLKDSV 317
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 425 ---EADTRSSSDPHLETTSTSTVSSMSTLSSSGELTDTHTSFADG---HTFLLEKPP 477
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
QY 318 LSLVHVKGRVLVALADGTLA----IFHRGEDGQWDLNHYHMLDLGHPHSIRCMAVVYDR 373
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 478 ---VPPKPKLKSPLGKGPVTFRDPPLLKQSSDSSELMAQQHHAASAG----- 519
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
QY 374 VWCYKKNKVHVIOPKTMOIEKSFDAHPRESQVRLAWIGDV-----WVSI 420
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 520 -----LASAGPARPRYLFQRRSKLW-GDPVESRGLPGPEDDKPTVIS 561
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
QY 421 RLDSTLRLYHAHTHQHQDQVD-----IEPYVSKMLGTGKLGFSFVRITALLVAGSRLWV 474
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 562 ELSSRLQQLNKDTRSLGEEPVGGLGSLLDPAKSPPIAARLFSSLGELSS--ISAQRSPG 619
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
QY 475 GTGNGVVISIPLTETVTLVHRGQLGLRANKTSPTSSEGEARPGGIHVYGGDSSDRAASSF 534
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 620 GPGGGASYSVRSGRYPVARRAPSPVKPASLERVEGLGAGAGAGRPFG-----L 669
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
QY 535 IPYCMAQAQLCFHGHRDAVKFEVSVPCNVLATLNGSVLDSPAEGPGPAAPASEVEGQKL 594
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 670 TPPTILKSSSLSIPHEPKVEVFRVRSVSARSFSPSPPLPSAPSGPGPGAPGRRPFQOK 729
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
QY 595 RNVLVLSGGEGYIDFRIGD-----GE-----DDETEGAGDMSQVKPVLKAEERSHII 642
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 730 PLQL-----WSKFDVGDWLESIHLEHRDRFEDHEIEGAH-----LPALTXXDDFVELG 777
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
QY 643 VWQVSY 648
   : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 778 VTRVGH 783

RESULT 7
SHK3_RAT STANDARD; PRT; 1815 AA.
AC Q9JLU4; Q9WUY7; Q9WV47;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE SH3 and multiple ankyrin repeat domains 3 (Shank3) (Proline-rich
DE synapse associated protein 2) (ProSAP2) (SPANK-2).
GN SHANK3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
```

```
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND INTERACTION WITH DLGAP1 AND DLG4.
RX MEDLINE=9458653; PubMed=10527873;
RA Boeckers T.M., Winter C., Smalla K.-H., Kreutz M.R., Bockmann J.,
RA Seidenbecher C., Garner C.C., Gundelfinger E.D.;
RT "Proline rich synapse associated proteins ProsAP1 and ProsAP2 interact
RL with synaptic proteins of the SAPAP/GKAP family.";
RL Biochem. Biophys. Res. Commun. 264:2476-2528(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH DLGAP1 AND CCTN.
RC STRAIN=Sprague-Dawley; TISSUE=Hippocampus;
RX MEDLINE=9360650; PubMed=10433268;
RA Naisbitt S., Kim E., Tu J.C., Xiao B., Sala C., Valtchanoff J.,
RA Weinberg R.J., Worley P.F., Sheng M.;
RT "Shank, a novel family of postsynaptic density proteins that binds to
RL the NMDA receptor/PSD-95/GKAP complex and cortactin.";
RL Neuron 23:569-582(1999).
RN [3]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Brain;
RX MEDLINE=20549637; PubMed=10958799;
RA Tobaben S., Suedhof T.C., Stahl B.;
RT "The G protein-coupled receptor CL1 interacts directly with proteins
RL of the shank family.";
RL J. Biol. Chem. 275:36204-36210(2000).
RN [4]
RP INTERACTION WITH HOMER PROTEINS; DLGAP1; MGLUR1A AND MGLUR5, AND
RC MUTAGENESIS LEU-1386 AND CYS-1389.
RX MEDLINE=9360651; PubMed=10433269;
RA Tu J.C., Xiao B., Naisbitt S., Yuan J.P., Petralia R.S., Brakeman P.,
RA Doan A., Aakalu V.K., Lanahan A.A., Sheng M., Worley P.F.;
RT "Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of
RL postsynaptic density proteins.";
RL Neuron 23:583-592(1999).
RN [5]
RP REVIEW.
RX MEDLINE=20267867; PubMed=10806096;
RA Sheng M., Kim E.;
RT "The Shank family of scaffold proteins.";
RL J. Cell Sci. 113:1851-1856(2000).
CC -!- FUNCTION: Seems to be an adapter protein in the postsynaptic
CC density (PSD) of excitatory synapses that interconnects receptors
CC of the postsynaptic membrane including NMDA-type and metabotropic
CC glutamate receptors via complexes with GKAP/PSD-95 and Homer,
CC respectively, and the actin-based cytoskeleton. May play a role in
CC the structural and functional organization of the dendritic spine
CC and synaptic junction.
CC -!- SUBUNIT: May homomultimerize via its SAM domain (By similarity).
CC Interacts with DLGAP1/GKAP, MGLUR1A and MGLUR5 C-termini via its
CC PDZ domain. Interacts with Homer-1, Homer-2, Homer-3 and
CC CCTN/cortactin SH3 domain. Is part of a complex with DLG4/PSD-95
CC and DLGAP1/GKAP.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of
CC neuronal cells. Extends into the region subjacent to the PSD.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=2; Synonyms=A;
CC IsoId=Q9JLU4-1; Sequence=Displayed;
CC Name=1;
CC IsoId=Q9JLU4-2; Sequence=VSP_006089;
CC Name=3;
CC IsoId=Q9JLU4-3; Sequence=VSP_006087, VSP_006088;
CC -!- TISSUE SPECIFICITY: Predominantly expressed in brain.
CC -!- SIMILARITY: BELONGS TO THE SHANK FAMILY.
CC -!- SIMILARITY: Contains 4 ANK repeats.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -!- CAUTION: Ref.2 sequence differs from that shown in positions 3 to
CC 37 due to a frameshift.
CC -!- CAUTION: Ref.3 sequence differs from that shown in positions 973
CC to 1000 due to a frameshift.
```

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; AJ133120; CAB45688.1; --  
DR EMBL; AF133301; AAF61375.1; ALT\_FRAME.  
DR EMBL; AF159047; AAD42976.1; ALT\_FRAME.  
DR InterPro; IPR002110; ANK.  
DR InterPro; IPR001478; PDZ.  
DR InterPro; IPR001660; SAM.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00023; ank; 5.  
DR Pfam; PF00595; PDZ; 1.  
DR Pfam; PF00536; SAM; 1.  
DR Pfam; PF00018; SH3; 1.  
DR ProDom; PD000066; SH3; 1.  
DR SMART; SM00248; ANK; 5.  
DR SMART; SM00228; PDZ; 1.  
DR SMART; SM00454; SAM; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS00088; ANK\_REPEAT; 4.  
DR PROSITE; PS0297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE; PS0106; PDZ; 1.  
DR PROSITE; PS0105; SAM\_DOMAIN; 1.  
DR PROSITE; PS00002; SH3; 1.  
KW ANK repeat; SH3 domain; SH3-binding; Coiled coil; Repeat;  
KW Alternative splicing.  
FT REPEAT 223 256 ANK 1.  
FT REPEAT 257 289 ANK 2.  
FT REPEAT 324 356 ANK 3.  
FT REPEAT 357 389 ANK 4.  
FT DOMAIN 472 488 COILED COIL (POTENTIAL).  
FT DOMAIN 545 604 SH3.  
FT DOMAIN 645 739 PDZ.  
FT SITE 1486 1492 SH3-BINDING (POTENTIAL).  
FT DOMAIN 1570 1590 COILED COIL (POTENTIAL).  
FT DOMAIN 1752 1815 SAM.  
FT DOMAIN 6 15 POLY-ALA.  
FT DOMAIN 895 901 POLY-PRO.  
FT DOMAIN 914 919 POLY-PRO.  
FT VARSPLIC 1204 1211 SPTPVHSP -> PRRRAGMV (in isoform 3).  
FT VARSPLIC 1212 1815 /FTid=VSP\_006087.  
FT VARSPLIC 1612 1620 /FTid=VSP\_006088.  
FT MUTAGEN 1386 1386 /FTid=VSP\_006089.  
FT MUTAGEN 1389 1389 P->L: ABOLISHES INTERACTION WITH HOMER-1  
FT MUTAGEN 1389 1389 F->C: ABOLISHES INTERACTION WITH HOMER-1  
FT CONFLICT 1 2 ISOFORM C.  
FT CONFLICT 32 32 MQ -> AA (IN REF. 2).  
FT CONFLICT 37 37 P -> Y (IN REF. 2).  
FT CONFLICT 68 68 R -> Q (IN REF. 2).  
FT CONFLICT 472 472 Q -> P (IN REF. 2).  
FT CONFLICT 781 781 H -> L (IN REF. 3).  
FT CONFLICT 902 907 R -> G (IN REF. 3).  
FT CONFLICT 912 917 YFDSG -> ILRLR (IN REF. 3).  
FT CONFLICT 921 921 FSPPPP -> SHGHQ (IN REF. 3).  
FT CONFLICT 963 963 R -> G (IN REF. 3).  
FT CONFLICT 1115 1115 S -> N (IN REF. 2).  
FT CONFLICT 1162 1162 S -> N (IN REF. 2).  
FT CONFLICT 1167 1167 G -> S (IN REF. 2).  
FT CONFLICT 1337 1337 S -> N (IN REF. 2).  
FT CONFLICT 1345 1345 G -> S (IN REF. 2).  
FT CONFLICT 1348 1348 S -> N (IN REF. 2).  
FT CONFLICT 1354 1354 E -> K (IN REF. 2).  
FT CONFLICT 1369 1369 S -> N (IN REF. 2).

FT CONFLICT 1507 1507 D -> G (IN REF. 2).  
FT CONFLICT 1715 1715 V -> A (IN REF. 2).  
SQ SEQUENCE 1815 AA; 193256 MW; 6632CF8F040766F9 CRC64;  
  
Query Match 4.2%; Score 146.5; DB 1; Length 1815;  
Best Local Similarity 21.3%; Pred. No. 0.31;  
Matches 147; Conservative 83; Mismatches 267; Indels 193; Gaps 30;  
  
QY 37 EDDAGNGVKA-----PGRDPLTCDREGDGEKPSAHTSPEKKAKELPEMDATSSR----- 87  
Db 1227 DSERGPLASPAFSPRSPAWIPVPAARAEAKPTREERKSPEDKSMILSVLDTSLQRPAGL 1286  
  
QY 88 -VMILTSTLTTSKVVIIDANQPGTVVDQFTVCNAHVLCISSIPAASDSYDPPGEMFLDS 146  
Db 1287 IVVHATSNQGEPRNLGAEEERPGTPELAPTPMQAAV--AEMPSPRAQPPGSI----- 1338  
  
QY 147 VNPEDPG-----ADGVLAGITLVGCATRCNVPRSNCSRGDTPVLDKQGEVATIA 197  
Db 1339 --PADPGPGQGSSEEEPELVFA-----VNLPPAQLSSSD-----EETREELARI- 1380  
  
QY 198 NGKVNPSQSTEEATEATEVDPDPGPSEPETATLRPGPLTEHVFTDPAPTSSGPGPGSENG 257  
Db 1381 -GLVPP---PEEFANGILLATPPPG-----PGPLPTTV-----PSPASG-KPSSELP 1422  
  
QY 258 PEPDSSS-----TRPEPEPSGDPTGAGSSAAPTMLGACNGWLY-VHSAVANWKKCL 308  
Db 1423 PAPERASDGSVEEADTRSSSDPHLETTSTISTVSSMSTLSSESGLTDTHTSFADG--- 1478  
  
QY 309 HSIKLDKSVLSLVHVKGRVLVALADGTLA----IFHRGEDGQWDLNHYHMLDLGHPHHSI 364  
Db 1479 HTFLEKPP---VPPKPKLKSPLGKGPVTFRDPPLLKQSSSEL-----MAQOHA- 1525  
  
QY 365 RCMAVVYDRVWCYKKNVHVIOPTMQIEKSFDAHPRESQVRLAWIGDV----- 416  
Db 1526 -----TSTGLTSAAGPARPRYLFORRSKLW-GDPVESRGLPGP 1562  
  
QY 417 -----WVSIRLDSTRLYHAHTHQHLDQVDVIEPYVSKMLGTGKLGFSFVITALLVAGSR 471  
Db 1563 EDDKPTVISELSSRLQNLNKDT---RSIGEEP-----VGGLGSLDDPAKKSPIAAAR 1611  
  
QY 472 LMVGTGNGVVISIPLTETVVLHRGQLLGURANKTSPTSCEGA---RPGGIHVYGDSS 527  
Db 1612 CAV-----VPSAGWLFSSSLGELSTISAQRSPPGGGASYSVRPSGRYPV----- 1656  
  
QY 528 DRAASFIPYCSMAQAQLCFHGHRAVKKFFVSVPGNVLATLNGSV----- 572  
Db 1657 ARRAPSVPKPAASLERVEGLGAGVGAGPPFGLTPPTTLKSSLSIPHEPKEVRFVRSVS 1716  
  
QY 573 -----LDSPAEGPGPAA-PASEVEGQKLRNVLSGGEGYIDFRIGDGED---DE 618  
Db 1717 ARSRSPSPPLSPSPSGSPSAGPRRPFQOKPLQWSKFDVGDWLESHLGEHRDRFEDH 1776  
  
QY 619 TEEGAGDMSQVKPVLKSAERSHIIVWQVSY 648  
Db 1777 EIEGAH-----LPALTKEDFVELGVTRVGH 1801

RESULT 8  
AMNH\_YEAST  
ID AMNH\_YEAST STANDARD; PRT; 1367 AA.  
AC P08640; P08068;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Glucoamylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase) (1,4-alpha-D-glucan glucohydrolase).  
GN STAI OR STA2 OR MAL5 OR YIR019C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.



FT REPEAT 172 179 13 (INCOMPLETE).

FT REPEAT 180 187 14 (INCOMPLETE).

FT DOMAIN 53 58 POLY-GLN.

FT DOMAIN 65 69 POLY-GLN.

FT DOMAIN 204 207 POLY-THR.

FT DOMAIN 208 220 POLY-SER.

FT DOMAIN 291 296 POLY-THR.

FT DOMAIN 304 307 POLY-THR.

FT DOMAIN 398 401 POLY-VAL.

FT CONFLICT 141 141 D -> DVPCDNPQPD (IN REF. 3).

FT CONFLICT 296 296 T -> A (IN REF. 2).

FT CONFLICT 441 441 S -> P (IN REF. 2).

FT CONFLICT 492 492 P -> S (IN REF. 2).

SQ SEQUENCE 634 AA; 65372 MW; F841347576BA6376 CRC64;

Query Match 4.0%; Score 137.5; DB 1; Length 634;

Best Local Similarity 23.8%; Pred. No. 0.31;

Matches 70; Conservative 35; Mismatches 98; Indels 91; Gaps 14;

QY 37 EDDAGNGVKPAPGRD-----PLTCDREGDGEKPSAHTSPEKKKAKELPE---MDATSS 86

Db 326 ESEVTTGVIVTSKDTIYTYCPLT-----ETTPVSTAP-----ATETPTGTVSTSEQ 374

QY 87 RWILT-----STLTSSKVI-----IDANQPGTVVDQFTVCNAHVL 123

Db 375 STTVITVTSBSSCTESEVTGGVVVTSEETVYVTFCLTENTPGT-----DSTPEASIP 430

QY 124 CISSIPAAADSDYPPGEMFLDSVNPEDPGADGVLGITLVGCATRCNVPRNSCSRGD 183

Db 431 PMETIPAGSESSMPAGE-----TSPAVPKSDVPATESA 463

QY 184 PVLD-KQGEVATIANGKVNPS--QSTEEATEATEVPDPGPSEPETATLRP---GPLTEH 237

Db 464 PVPEMTPAGSOPSIPAGETSPAVPKSDVPATESAPAPEMTPAGTETKPAAPKSSAPATE- 522

QY 238 VFTDPAP-----TPSSGPQPGSENGPEPDSSTRPEP-EPSCGPTGAGSSAAPT 286

Db 523 -----PSPVAPGTESAPAGPGAGSSPKSSVLASETSPIAPGAETAPAGSSGAI 572

RESULT 10

MINT MOUSE

ID MINT MOUSE STANDARD; PRT; 3644 AA.

AC Q62504; Q80TN9; Q99PS4; Q9QZW2;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Mx2-interacting protein (SMART/HDAC1 associated repressor protein).

GN MINT OR SHARP OR KIAA0929.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE OF 49-3644 FROM N.A. (ISOFORM 1), FUNCTION, SUBCELLULAR

RP LOCATION, TISSUE SPECIFICITY, DNA-BINDING, AND INTERACTION WITH MSX2.

RC TISSUE-Testis;

RX MEDLINE=99379811; PubMed=10451362;

RA Newberry E.P., Latifi T., Towler D.A.;

RT "The RRM domain of MINT, a novel mxx2 binding protein, recognizes and

RT regulates the rat osteocalcin promoter.";

RL Biochemistry 38:10678-10690(1999).

RN [2]

RP SEQUENCE OF 1-112 FROM N.A.

RA Adachi J., Aizawa K., Akimura T., Arakawa T., Carninci P., Fukuda S.,

RA Hashizume W., Hayashida K., Hirozane T., Hori F., Imotani K.,

RA Ishii Y., Itoh M., Kagawa I., Kawai J., Kojima Y., Kondo S., Konno H.,

RA Koya S., Miyazaki A., Murata M., Nakamura M., Nomura K., Numazaki R.,

RA Ohno M., Ohsato N., Saito R., Sakazume N., Sano H., Sasaki D.,

RA Sato K., Shibata K., Shiraki T., Tagami M., Takeda Y., Waki K.,

RA Watahiki A., Muramatsu M., Hayashizaki Y.;

RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 69-3644 FROM N.A. (ISOFORM 2), AND VARIANTS THR-348;

RP PHE-762; PHE-773 AND LEU-933.

RC STRAIN=ICR; TISSUE=Brain;

RA Sakamoto T., Gotou T., Isagawa Y., Mimura H., Kimura K., Kawaichi M.;

RT "MINT/spen negatively regulates Notch signaling by inhibiting RBP-

RT J/Su(H) activity.";

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE OF 318-578 FROM N.A.

RC TISSUE=Cochlea;

RX MEDLINE=97237053; PubMed=9119401;

RA Crozet F., El-Amraoui A., Blanchard S., Lenoir M., Ripoll C., Vago P.,

RA Hamel C., Fizames C., Levi-Acobas F., Depetris D., Mattei M.-G.,

RA Weil D., Pujol R., Petit C.;

RT "Cloning of the genes encoding two murine and human cochlear

RT unconventional type I myosins.";

RL Genomics 40:332-341(1997).

RN [5]

RP SEQUENCE OF 2598-3644 FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=22579291; PubMed=12693553;

RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,

RA Nakajima D., Nagase T., Ohara O., Koga H.;

RT "Prediction of the coding sequences of mouse homologues of KIAA gene:

RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous

RT cDNAs identified by screening of terminal sequences of cDNA clones

RT randomly sampled from size-fractionated libraries.";

RL DNA Res. 10:35-48(2003).

RN [6]

RP TISSUE SPECIFICITY.

RX MEDLINE=22261914; PubMed=12374742;

RA Oswald F., Kostezka U., Astrahantseff K., Bourteele S., Dillinger K.,

RA Zechner U., Ludwig L., Wilda M., Hameister H., Knoechel W., Liptay S.,

RA Schmid R.M.;

RT "SHARP is a novel component of the Notch/RBP-Jkappa signalling

RT pathway.";

RL EMBO J. 21:5417-5426(2002).

RN [7]

RP FUNCTION, AND TISSUE SPECIFICITY.

RX MEDLINE=22483652; PubMed=12594956;

RA Kuroda K., Han H., Tani S., Tanigaki K., Tun T., Furukawa T.,

RA Taniguchi Y., Kurooka H., Hamada Y., Toyokuni S., Honjo T.;

RT "Regulation of marginal zone B cell development by MINT, a suppressor

RT of Notch/RBP-J signaling pathway.";

RL Immunity 18:301-312(2003).

CC -!- FUNCTION: Essential corepressor protein, which probably regulates

CC different key pathways such as the Notch pathway. Negative

CC regulator of the Notch pathway via its interaction with RBPSUH,

CC which prevents the association between NOTCH1 and RBPSUH, and

CC therefore suppresses the transactivation activity of Notch

CC signaling. Blocks the differentiation of precursor B cells into

CC marginal zone B cells. Probably represses transcription via the

CC recruitment of large complexes containing histone deacetylase

CC proteins. May bind both to DNA and RNA.

CC -!- SUBUNIT: Interacts with NCOR2, HDAC1, HDAC2, RBBP4, MBD3 and

CC MTA1L1. Interacts with the nuclear receptors RAR and PPARG.

CC Interacts with RAR in absence of ligand. Bind to the steroid

CC receptor RNA coactivator SRA (By similarity). Interacts with MSX2.

CC Interacts with RBPSUH; this interaction may prevent the

CC interaction between RBPSUH and NOTCH1.

CC -!- SUBCELLULAR LOCATION: Nuclear. Associates with chromatin.

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Name=1;

CC IsoId=Q62504-1; Sequence=Displayed;

CC Name=2;

CC IsoId=Q62504-2; Sequence=VSP\_008564;

CC Note=No experimental confirmation available;

CC -!- TISSUE SPECIFICITY: Highly expressed in testis. Expressed at lower

CC level in brain, lung, spleen, liver and kidney. Weakly expressed

CC in cardiac and skeletal muscles and ovary. In spleen, it is

CC expressed in follicular B-cells, while it is weakly expressed in

CC marginal zone B-cells.



[3]  
RP SEQUENCE OF 642-1115 FROM N.A. (ISOFORM N-CAM 180).  
RX MEDLINE=88283628; PubMed=3396534;  
RA Barbas J.A., Chaix J.C., Steinmetz M., Goridis C.;  
RT "Differential splicing and alternative polyadenylation generates  
RL distinct NCAM transcripts and proteins in the mouse.";  
EMBO J. 7:625-632(1988).  
[4]  
RP SEQUENCE OF 804-1081 FROM N.A. (ISOFORM N-CAM 180).  
RC STRAIN=C57BL/6J; TISSUE=Brain;  
RX MEDLINE=88247737; PubMed=2454455;  
RA Barthels D., Vopper G., Wille W.;  
RT "NCAM-180, the large isoform of the neural cell adhesion molecule of  
the mouse, is encoded by an alternatively spliced transcript.";  
RL Nucleic Acids Res. 16:4217-4225(1988).  
[5]  
RP SEQUENCE OF 702-1115 FROM N.A. (ISOFORM N-CAM 140).  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RX MEDLINE=89251563; PubMed=2721486;  
RA Santoni M.J., Barthels D., Vopper G., Boned A., Goridis C., Wille M.;  
RT "Differential exon usage involving an unusual splicing mechanism  
generates at least eight types of NCAM cDNA in mouse brain.";  
RL EMBO J. 8:385-392(1989).  
[6]  
RP SEQUENCE OF 20-36.  
RX MEDLINE=86140120; PubMed=3512556;  
RA Rougon G., Marshak D.R.;  
RT "Structural and immunological characterization of the amino-terminal  
domain of mammalian neural cell adhesion molecules.";  
RL J. Biol. Chem. 261:3396-3401(1986).  
CC -!- FUNCTION: This protein is a cell adhesion molecule involved in  
neuron-neuron adhesion, neurite fasciculation, outgrowth of  
neurons, etc.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;  
CC Name=N-CAM 180;  
CC IsoId=Pl3595-1; Sequence=Displayed;  
CC Name=N-CAM 140;  
CC IsoId=Pl3595-2; Sequence=VSP\_002588;  
CC Name=N-CAM 120;  
CC IsoId=Pl3594-1; Sequence=External;  
CC -!- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.  
CC -!- SIMILARITY: Contains 2 fibronectin type III domains.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X07200; CAA30177.1; -;  
DR EMBL; Y00051; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; X06328; CAA29641.1; -;  
DR EMBL; X07195; CAA30173.1; -;  
DR EMBL; X07244; CAA30230.1; -;  
DR EMBL; X15051; CAA33150.1; -;  
DR EMBL; X15052; CAA33151.1; -;  
DR PIR; A29673; IJMSNL.  
DR MGD; MGI:97281; Ncam1.  
DR InterPro; IPR008957; FN\_III-like.  
DR InterPro; IPR003961; FN\_III.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003598; Ig\_c2.  
DR Pfam; PF00041; fn3; 2.  
DR Pfam; PF00047; ig; 5.  
DR SMART; SM00060; FN3; 2.  
DR SMART; SM00408; Igc2; 5.  
DR PROSITE; PS50835; IG\_LIKE; 5.  
KW Cell adhesion; Glycoprotein; Transmembrane; Repeat;  
Immunoglobulin domain; Alternative splicing; Signal; Heparin-binding.

FT SIGNAL 1 19  
FT CHAIN 20 1115  
FT NEURAL CELL ADHESION MOLECULE 1, 180 kDa  
FT ISOFORM.  
FT EXTRACELLULAR (POTENTIAL).  
FT POTENTIAL.  
FT CYTOPLASMIC (POTENTIAL).  
FT IG-LIKE C2-TYPE 1.  
FT IG-LIKE C2-TYPE 2.  
FT IG-LIKE C2-TYPE 3.  
FT IG-LIKE C2-TYPE 4.  
FT IG-LIKE C2-TYPE 5.  
FT FIBRONECTIN TYPE-III 1.  
FT FIBRONECTIN TYPE-III 2.  
FT HEPARIN-BINDING (POTENTIAL).  
FT HEPARIN-BINDING (POTENTIAL).  
FT PROBABLE.  
FT PROBABLE.  
FT PROBABLE.  
FT PROBABLE.  
FT PROBABLE.  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT Missing (in isoform N-CAM 140).  
FT /FTId=VSP\_002588.  
FT SQ SEQUENCE 1115 AA; 119351 MW; 2C93DCD474CFBCAF CRC64;  
Query Match 3.8%; Score 132.5; DB 1; Length 1115;  
Best Local Similarity 22.5%; Pred. No. 1.2;  
Matches 83; Conservative 40; Mismatches 127; Indels 119; Gaps 17;  
QY 22 LWCAAGVNLGWRPNEDDAGNVKPAQGRD-----PL-----TCDREGDGE 62  
Db 742 LMCIA-VNLCG-----KAGPGAK--GKDMEEGKAFAKDESKPIVEVTEERTPNH 791  
QY 63 PKSAHTSPEKKKAKELPEM--DATSSRVWLTS--TLTTSKVVIIDA-----NQP----- 108  
Db 792 DGGKHTEPNETTPTLPELPADTTATVEDMLPSVTTVTNDSITITETATAQNSPTSETT 851  
QY 109 -----GTVVDQFTV-----CNAHVLCISSIPAASDSDYPPGEMFLDSVNPEDP 152  
Db 852 TLTSSIAAPPATTVPDSNSVPAGQATPSKGVTSASSSPASAPKVAFLVDL---SDTPTSAP 908  
QY 153 GAD-----GVLAGITLVGCATRCNVPRNSCNRGDTPLVDKQGQEVATIANGKV 201  
Db 909 SASNLSSTVLANQGAVALSPSPASAGETSKAPPASKASPAPTPTPAGASPLAAVAAPAT 968  
QY 202 NPSQSTBEATEATEVPDPGPSE-----PETATLRPGPLTEHVFTDP----- 242  
Db 969 DAPQAKQEA-PSITKGPDPPEPTQPGTVKNPPEAATAPAPSKSKAATTNPSQGEDLKMDEGN 1027  
QY 243 -----APTPSSGPQ-PGSENGPEPSSSTRPEP-----EPGSDPT 276  
Db 1028 FKTPDIDLAKDVFAALGSPRPATGASQASELAPSPADSAPVPPAPAKTEKGPVETKSEPP 1087  
QY 277 GAGSSAAPT 285  
Db 1088 ESEAKEAPT 1096  
RESULT 12  
OGFR\_HUMAN  
ID OGFR\_HUMAN STANDARD; PRT; 677 AA.  
AC Q9NZT2; Q96029; Q96CM2; Q9BQW1; Q9H4H0; Q9H7J5; Q9NZT3; Q9NZT4;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Opioid growth factor receptor (OGFr) (Zeta-type opioid receptor) (7-60  
DE protein).  
GN OGFR.

|    |  |  |  |
|----|--|--|--|
| OS | Homo sapiens (Human).  |  |  |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;      |  |  |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.             |  |  |
| OX | NCBI_TaxID=9606;   |  |  |
| RN | [1]  |  |  |
| RP | SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).                                 |  |  |
| RC | TISSUE=Placenta;   |  |  |
| RX | MEDLINE=20143540; PubMed=10677613;                                     |  |  |
| RA | Zagon I.S., Verderame M.F., Allen S.S., McLaughlin P.J.;               |  |  |
| RT | "Cloning, sequencing, chromosomal location, and function of cDNAs      |  |  |
| RT | encoding an opioid growth factor receptor (OGFr) in humans.";          |  |  |
| RL | Brain Res. 856:75-83(2000).  |  |  |
| RN | [2]  |  |  |
| RP | SEQUENCE FROM N.A.   |  |  |
| RA | Takanosu M., Liu J., Mayne R., Wood B.M., Brewton R.G.;                |  |  |
| RT | "Genomic structure of human gene 7-60.";                               |  |  |
| RL | Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.                |  |  |
| RN | [3]  |  |  |
| RP | SEQUENCE FROM N.A.   |  |  |
| RX | MEDLINE=21638749; PubMed=11780052;                                     |  |  |
| RA | Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,     |  |  |
| RA | Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,     |  |  |
| RA | Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,            |  |  |
| RA | Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,      |  |  |
| RA | Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,            |  |  |
| RA | Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,   |  |  |
| RA | Clegg S., Copley V.E., Collier R.E., Connor R.E., Corby N.R.,          |  |  |
| RA | Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,             |  |  |
| RA | Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,       |  |  |
| RA | Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  |  |  |
| RA | Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  |  |  |
| RA | Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., |  |  |
| RA | Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  |  |  |
| RA | Lehvaslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,    |  |  |
| RA | Marsh V.L., Martin S.L., McConnachie L.J., McIlroy K., McMurray A.A.,  |  |  |
| RA | Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,      |  |  |
| RA | Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,              |  |  |
| RA | Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,         |  |  |
| RA | Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,    |  |  |
| RA | Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,      |  |  |
| RA | Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,    |  |  |
| RA | Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,              |  |  |
| RA | Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A., |  |  |
| RA | Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S., |  |  |
| RA | Rogers J.;   |  |  |
| RT | "The DNA sequence and comparative analysis of human chromosome 20.";   |  |  |
| RL | Nature 414:865-871(2001).  |  |  |
| RN | [4]  |  |  |
| RP | SEQUENCE FROM N.A. (ISOFORM 2).  |  |  |
| RC | TISSUE=Skin;   |  |  |
| RX | MEDLINE=22388257; PubMed=12477932;                                     |  |  |
| RA | Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,               |  |  |
| RA | Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,    |  |  |
| RA | Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,      |  |  |
| RA | Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,        |  |  |
| RA | Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,          |  |  |
| RA | Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  |  |  |
| RA | Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,     |  |  |
| RA | Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  |  |  |
| RA | Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,    |  |  |
| RA | Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  |  |  |
| RA | Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,          |  |  |
| RA | Fahy J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,   |  |  |
| RA | Whiting M., Madan A., Young A.C., Green E.D., Dickson M.C.,            |  |  |
| RA | Blakesley R.W., Touchman J.W., Schmutz J., Myers R.M.,                 |  |  |
| RA | Rodriguez A.C., Grimwood J., Schmutz J., Skalska U., Smailus D.E.,     |  |  |
| RA | Butterfield Y.S.N., Krzywinski M.I., Marra M.A.;                       |  |  |
| RA | Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;                    |  |  |
| RT | "Generation and initial analysis of more than 15,000 full-length       |  |  |
| RT | human and mouse cDNA sequences.";                                      |  |  |
| RL | Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).                    |  |  |
| RN | [5]  |  |  |
| RP | SEQUENCE OF 37-677 FROM N.A. (ISOFORM 1).                              |  |  |

|    |   |  |  |
|----|---|--|--|
| RC | TISSUE=Spleen;  |  |  |
| RX | MEDLINE=21082933; PubMed=11214971;  |  |  |
| RA | Hattori A., Okumura K., Nagase T., Kikuno R., Hirose M., Ohara O.;  |  |  |
| RT | "Characterization of long cDNA clones from human adult spleen.";  |  |  |
| RL | DNA Res. 7:357-366(2000).   |  |  |
| RN | [6]   |  |  |
| RP | REVIEW.   |  |  |
| RX | MEDLINE=21888341; PubMed=11890982;  |  |  |
| RA | Zagon I.S., Verderame M.F., McLaughlin P.J.;  |  |  |
| RT | "The biology of the opioid growth factor receptor (OGFr).";   |  |  |
| RL | Brain Res. Brain Res. Rev. 38:351-376(2002).  |  |  |
| CC | -!- FUNCTION: Receptor for opioid growth factor (OGF), also known as  |  |  |
| CC | Met-enkephalin. Seems to be involved in growth regulation.  |  |  |
| CC | -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity). The OGF/OGFR   |  |  |
| CC | complex is probably translocated to the nucleus.  |  |  |
| CC | -!- ALTERNATIVE PRODUCTS:   |  |  |
| CC | Event=Alternative splicing; Named isoforms=2;   |  |  |
| CC | Name=1;   |  |  |
| CC | Isoid=Q9NZT2-1; Sequence=Displayed;   |  |  |
| CC | Name=2;   |  |  |
| CC | Isoid=Q9NZT2-2; Sequence=VSP_004060;  |  |  |
| CC | Note=No experimental confirmation available;  |  |  |
| CC | -!- TISSUE SPECIFICITY: Highly expressed in the heart and liver,  |  |  |
| CC | moderately in skeletal muscle and kidney and to a lesser extent in  |  |  |
| CC | brain and pancreas. Expressed in fetal tissues including liver and  |  |  |
| CC | kidney.   |  |  |
| CC | -!- CAUTION: Ref.2 sequence differs from that shown due to frameshifts  |  |  |
| CC | in position 1 and 181.  |  |  |
| CC | -!- CAUTION: Ref.3 (CAC28882) sequence differs from that shown due to   |  |  |
| CC | erroneous gene model prediction.  |  |  |
| CC | -!- CAUTION: Ref.3 (CAC12749) sequence differs from that shown due to   |  |  |
| CC | a frameshift in position 514.   |  |  |
| CC | -----   |  |  |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration  |  |  |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation -   |  |  |
| CC | the European Bioinformatics Institute. There are no restrictions on its   |  |  |
| CC | use by non-profit institutions as long as its content is in no way  |  |  |
| CC | modified and this statement is not removed. Usage by and for commercial   |  |  |
| CC | entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ). |  |  |
| CC | -----   |  |  |
| DR | EMBL; AF172451; AAF64404.1; -   |  |  |
| DR | EMBL; AF172452; AAF64405.1; -   |  |  |
| DR | EMBL; AF172453; AAF64406.1; -   |  |  |
| DR | EMBL; AF112980; AAD03745.1; ALT FRAME.  |  |  |
| DR | EMBL; AF109134; AAD03737.1; ALT FRAME.  |  |  |
| DR | EMBL; AL035669; CAC28882.1; ALT SEQ.  |  |  |
| DR | EMBL; AL035669; CAC12749.1; ALT_FRAME.  |  |  |
| DR | EMBL; BC014137; AAHL14137.1; -  |  |  |
| DR | EMBL; AK024485; BAB15775.1; -   |  |  |
| DR | GeneW; HGNC:15768; OGFR.  |  |  |
| DR | MIM; 606459; -  |  |  |
| DR | GO; GO:0004985; F:opioid receptor activity; NAS.  |  |  |
| DR | GO; GO:0001558; P:regulation of cell growth; NAS.   |  |  |
| DR | InterPro; IPR006757; OGFR_N.  |  |  |
| DR | InterPro; IPR006770; OGFR_repeat.   |  |  |
| DR | Pfam; PF04664; OGFR_N; 1.   |  |  |
| DR | Pfam; PF04680; OGFR_repeat; 7.  |  |  |
| KW | Receptor; Growth regulation; Repeat; Alternative splicing.  |  |  |
| FT | DOMAIN 267 283 BIPARTITE NUCLEAR LOCALIZATION SIGNAL  |  |  |
| FT | (POTENTIAL).  |  |  |
| FT | 7 X APPROXIMATE 20 AA TANDEM REPEATS.   |  |  |
| FT | DOMAIN 517 656  |  |  |
| FT | REPEAT 517 536  |  |  |
| FT | REPEAT 537 556  |  |  |
| FT | REPEAT 557 576  |  |  |
| FT | REPEAT 577 596  |  |  |
| FT | REPEAT 597 616  |  |  |
| FT | REPEAT 617 636  |  |  |
| FT | REPEAT 637 656  |  |  |
| FT | VARSPLIC 558 577 Missing (in isoform 2).  |  |  |
| FT | FT/ID=VSP_004060.   |  |  |
| FT | DG -> EA (IN REF. 2).   |  |  |
| FT | DAGDEDESEFPRAARPSFQSRMTGSRNWRATRDMCRYRH   |  |  |
| FT | CONFLICT 27 28  |  |  |
| FT | CONFLICT 37 80  |  |  |









GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2004, 11:05:23 ; Search time 128 Seconds  
(without alignments)  
1604.706 Million cell updates/sec

Title: US-10-019-495-9  
Perfect score: 3473  
Sequence: 1 MKNVPVPVYCRPLVEKPTM.....VLSKAERSHIIWQVSYTPE 651

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL 25:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archheap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID     | Description        |
|------------|--------|-------------|--------|-----------|--------------------|
| 1          | 3191   | 91.9        | 1328   | 11 Q8CHB5 | Q8chb5 mus musculu |
| 2          | 3057   | 88.0        | 599    | 6 Q95LW5  | Q95lw5 macaca fasc |
| 3          | 2870   | 82.6        | 587    | 11 Q8CAB6 | Q8cab6 mus musculu |
| 4          | 1892   | 54.5        | 827    | 11 Q8C7W0 | Q8c7w0 mus musculu |
| 5          | 1821.5 | 52.4        | 1307   | 11 Q8CJC2 | Q8cjc2 mus musculu |
| 6          | 1814.5 | 52.2        | 1179   | 4 O60271  | O60271 homo sapien |
| 7          | 1814.5 | 52.2        | 1311   | 4 Q8IZX7  | Q8izx7 homo sapien |
| 8          | 943    | 27.2        | 734    | 4 Q9H811  | Q9h811 homo sapien |
| 9          | 623.5  | 18.0        | 200    | 4 Q7Z5Q8  | Q7z5q8 homo sapien |
| 10         | 460.5  | 13.3        | 766    | 4 O60905  | O60905 homo sapien |
| 11         | 460.5  | 13.3        | 780    | 4 Q86WC8  | Q86wc8 homo sapien |
| 12         | 449    | 12.9        | 1082   | 11 Q80U35 | Q80u35 mus musculu |
| 13         | 442.5  | 12.7        | 1510   | 4 Q9Y4G3  | Q9y4g3 homo sapien |
| 14         | 442.5  | 12.7        | 2063   | 4 Q96PE2  | Q96pe2 homo sapien |
| 15         | 436.5  | 12.6        | 1101   | 5 Q9VEF9  | Q9vef9 drosophila  |
| 16         | 389.5  | 11.2        | 760    | 6 Q7YRB3  | Q7yrb3 papio hamad |

|    |       |     |      |           |                     |
|----|-------|-----|------|-----------|---------------------|
| 17 | 333.5 | 9.6 | 1284 | 4 Q9HCE6  | Q9hce6 homo sapien  |
| 18 | 320   | 9.2 | 720  | 11 Q8VDH5 | Q8vdh5 mus musculu  |
| 19 | 320   | 9.2 | 1241 | 11 Q8C1A1 | Q8clal mus musculu  |
| 20 | 306   | 8.8 | 988  | 4 O14665  | O14665 homo sapien  |
| 21 | 305.5 | 8.8 | 1405 | 4 O15013  | O15013 homo sapien  |
| 22 | 291.5 | 8.4 | 701  | 11 Q80VH8 | Q80vh8 mus musculu  |
| 23 | 291.5 | 8.4 | 1139 | 11 Q8C033 | Q8c033 mus musculu  |
| 24 | 265   | 7.6 | 1458 | 5 Q9VX48  | Q9vx48 drosophila   |
| 25 | 265   | 7.6 | 1740 | 5 Q86B45  | Q86b45 drosophila   |
| 26 | 254.5 | 7.3 | 755  | 5 Q81GL6  | Q8igl6 drosophila   |
| 27 | 233   | 6.7 | 596  | 4 Q9NVT3  | Q9nvt3 homo sapien  |
| 28 | 210   | 6.0 | 193  | 11 Q8BZY4 | Q8bzy4 mus musculu  |
| 29 | 146.5 | 4.2 | 249  | 16 Q9RDN7 | Q9rdn7 streptomyce  |
| 30 | 146.5 | 4.2 | 985  | 12 Q67643 | Q67643 gallid herp  |
| 31 | 145   | 4.2 | 483  | 12 Q83344 | Q83344 murid herpe  |
| 32 | 144.5 | 4.2 | 1054 | 11 Q9JM99 | Q9jnm99 mus musculu |
| 33 | 144.5 | 4.2 | 1527 | 3 Q8WZL5  | Q8wzl5 yarrowia li  |
| 34 | 144.5 | 4.2 | 2344 | 5 Q9N3Y8  | Q9n3y8 caenorhabdi  |
| 35 | 142   | 4.1 | 300  | 10 Q07373 | Q07373 chlamydomon  |
| 36 | 141.5 | 4.1 | 582  | 16 Q82F59 | Q82f59 streptomyce  |
| 37 | 140   | 4.0 | 483  | 12 Q9DYE3 | Q9dye3 murine herp  |
| 38 | 140   | 4.0 | 647  | 10 Q8S148 | Q8s148 oryza sativ  |
| 39 | 138.5 | 4.0 | 464  | 16 Q8FS16 | Q8fs16 corynebacte  |
| 40 | 138.5 | 4.0 | 667  | 16 Q9RJY5 | Q9rjy5 streptomyce  |
| 41 | 138.5 | 4.0 | 825  | 5 Q7YZW4  | Q7yzw4 caenorhabdi  |
| 42 | 138.5 | 4.0 | 1032 | 5 P91365  | P91365 caenorhabdi  |
| 43 | 138.5 | 4.0 | 2232 | 5 Q81FX6  | Q81fx6 caenorhabdi  |
| 44 | 138   | 4.0 | 586  | 16 Q98J75 | Q98j75 rhizobium l  |
| 45 | 138   | 4.0 | 1013 | 16 Q9KE85 | Q9ke85 bacillus ha  |

ALIGNMENTS

RESULT 1  
Q8CHB5  
ID Q8CHB5 PRELIMINARY; PRT; 1328 AA.  
AC Q8CHB5;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE MKIAA1066 protein (Fragment).  
GN MAPK8IP3 OR MKIAA1066.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Hara Y., Nagase T.,  
RA Ohara O., Koga H.;  
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:  
RT I. The complete nucleotide sequences of 100 mouse KIAA-homologous  
RT cDNAs identified by screening of terminal sequences of cDNA clones  
RT randomly sampled from size-fractionated libraries.";  
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB093282; BAC41466.2; -.  
DR MGD; MGI:1353598; Mapk8ip3.  
DR GO; GO:0005737; C:cytoplasm; IDA.  
DR GO; GO:0005078; F:MAP-kinase scaffold activity; IDA.  
DR GO; GO:0007254; P:JNK cascade; IDA.  
FT NON\_TER 1  
SQ SEQUENCE 1328 AA; 146328 MW; 647EF78FC762B29F CRC64;

Query Match 91.9%; Score 3191; DB 11; Length 1328;  
Best Local Similarity 91.2%; Pred. No. 7e-223;  
Matches 594; Conservative 24; Mismatches 33; Indels 0; Gaps 0;

Qy 1 MKNVPVPVYCRPLVEKPTMKLWCAAGVNLSCWRPNEDDAGNGVKPAPGRDPLTCDREGD 60  
Db 678 MKNVPVPVYCRPLVEKDPSTKLWCAAGVNLSCWKPHEDSSNGKPKVPGRDPLTCDREGE 737

QY 61 GEPKSAHTSPEKKKAKELPEMDATSSRVWILTSTLTTSKVVIIDANQPVTVDQFTVCNA 120  
Db 738 GEPKSTHPSPEKKKAKETPEADATSSRVWILTSTLTTSKVVIIDANQPVTVDQFTVCNA 797  
QY 121 HVLCISSIPAASDSYPPGEMFLDSVNPEDPGADGVLAGITLVGCATRCNVPRSNCSRR 180  
Db 798 HVLCISSIPAASDSYPPGEMFLDSVNPEDPGADGVLAGITLVGCATRCNVPRSNCSRR 857  
QY 181 GDTPLVDKGOGEVATIANGKVNPSQSTEEATEATEVDPGPSPETATLRPGPLTEHVFT 240  
Db 858 GDTPLVDKGOGEVATTANGKVNPSQSTEEATEATEVDPGPSPSEATTVRPGPLTEHVFT 917  
QY 241 DPAPTPSSGPOGSENGPEPDSSTRPEPEPSGDPTGAGSSAAPTMMWLGAGQNGWLYVHSA 300  
Db 918 DPAPTPSSSTQPASENGSENGTIVQPQVEPSGELSTTTSSAAPTMMWLGAGQNGWLYVHSA 977  
QY 301 VANWKKCLHSIKLKDSVLSLVHVKGRLVALADGTLAIFHRGEDGQWDLSNYHLMDLGHP 360  
Db 978 VANWKKCLHSIKLKDSVLSLVHVKGRLVALADGTLAIFHRGEDGQWDLSNYHLMDLGHP 1037  
QY 361 HHSIRCMVVDVRCVWCGYKKNKHVVIQPKTMQIEKSFDAHPRRESQVRQLAWIGDVWVSI 420  
Db 1038 HHSIRCMVVDVRCVWCGYKKNKHVVIQPKTMQIEKSFDAHPRRESQVRQLAWIGDVWVSI 1097  
QY 421 RLDSTLRLYHAHTHQHLDQVDIEPYVSKMLGTGKLGFSFVRITALLVAGSRLWVGTVNGV 480  
Db 1098 RLDSTLRLYHAHTHQHLDQVDIEPYVSKMLGTGKLGFSFVRITALLVAGSRLWVGTVNGV 1157  
QY 481 VISIPLTETVVLHRGQLLGLRANKTSPTSGEGARPGGIIHVYGDSSDRAASSFIPYCSM 540  
Db 1158 VISIPLTETVVLHRGQLLGLRANKTSPTSGEGTRPGGIIHVYGDSSDRAASSFIPYCSM 1217  
QY 541 AQAQLCFHGHHRDAVKFFVSVPGNVLATLNGSVLDSPAEGPAPASEVEGQKLRNLVL 600  
Db 1218 AQAQLCFHGHHRDAVKFFVSVPGNVLATLNGSVLDSPAEGPAPADAEGQKLNALVL 1277  
QY 601 SGGEYIDFRIGDEDDTEEGAGDMSQVKPVLKSAERSHIIWQVSYTPE 651  
Db 1278 SGGEYIDFRIGDEDDTEECAGDVNQTKPSLSKAERSHIIWQVSYTPE 1328

RESULT 2  
Q95LW5  
ID Q95LW5 PRELIMINARY; PRT; 599 AA.  
AC Q95LW5;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;  
OC Cercopithecinae; Macaca.  
OX NCBI\_TaxID=9541;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,  
RA Terao K., Sugano S.;  
RT "Isolation of novel full-length cDNA clones from macaque testis cDNA  
libraries."  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB071076; BAB64469.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 599 AA; 64132 MW; 5650634938EC193 CRC64;

Query Match 88.0%; Score 3057; DB 6; Length 599;  
Best Local Similarity 98.3%; Pred. No. 1.3e-213;  
Matches 571; Conservative 3; Mismatches 7; Indels 0; Gaps 0;  
QY 1 MKNVPPVYCRPLVEKDPMTKMLWCAAGVNLGWRPNEDDAGNGVKPAPGRDPLTCDREGD 60  
Db 1 MKNVPPVYCRPLVEKDPMTKMLWCAAGVNLGSKPSEDDGGNGVKPAPGRDPLTCDREGD 60

QY 61 GEPKSAHTSPEKKKAKELPEMDATSSRVWILTSTLTTSKVVIIDANQPVTVDQFTVCNA 120  
Db 61 GEPKSAHGSPEKKKVKELPEMDATSSRVWILTSTLTTSKVVIIDANQPVTVDQFTVCNA 120  
QY 121 HVLCISSIPAASDSYPPGEMFLDSVNPEDPGADGVLAGITLVGCATRCNVPRSNCSRR 180  
Db 121 HVLCISSIPAASDSYPPGEMFLDSVNPEDPGADGVLAGITLVGCATRCNVPRSNCSRR 180  
QY 181 GDTPLVDKGOGEVATIANGKVNPSQSTEEATEATEVDPGPSPETATLRPGPLTEHVFT 240  
Db 181 GDTPLVDKGOGEVATIANGKVNPSQSTEEATEATEVDPGPSPETATLRPGPLTEHVFT 240  
QY 241 DPAPTPSSGPOGSENGPEPDSSTRPEPEPSGDPTGAGSSAAPTMMWLGAGQNGWLYVHSA 300  
Db 241 DPAPTPSSGPOGSENGPEPDSSTQPEPEPSGDPTGAGSSAAPTMMWLGAGQNGWLYVHSA 300  
QY 301 VANWKKCLHSIKLKDSVLSLVHVKGRLVALADGTLAIFHRGEDGQWDLSNYHLMDLGHP 360  
Db 301 VANWKKCLHSIKLKDSVLSLVHVKGRLVALADGTLAIFHRGEDGQWDLSNYHLMDLGHP 360  
QY 361 HHSIRCMVVDVRCVWCGYKKNKHVVIQPKTMQIEKSFDAHPRRESQVRQLAWIGDVWVSI 420  
Db 361 HHSIRCMVVDVRCVWCGYKKNKHVVIQPKTMQIEKSFDAHPRRESQVRQLAWIGDVWVSI 420  
QY 421 RLDSTLRLYHAHTHQHLDQVDIEPYVSKMLGTGKLGFSFVRITALLVAGSRLWVGTVNGV 480  
Db 421 RLDSTLRLYHAHTHQHLDQVDIEPYVSKMLGTGKLGFSFVRITALLVAGSRLWVGTVNGV 480  
QY 481 VISIPLTETVVLHRGQLLGLRANKTSPTSGEGARPGGIIHVYGDSSDRAASSFIPYCSM 540  
Db 481 VISIPLTETVVLHRGQLLGLRANKTSPTSGEGARPGGIIHVYGDSSDRAASSFIPYCSM 540  
QY 541 AQAQLCFHGHHRDAVKFFVSVPGNVLATLNGSVLDSPAEGPG 581  
Db 541 AQAQLCFHGHHRDAVKFFVSVPGNVLATLNGSVLDSPAEPGLG 581

RESULT 3  
Q8CAB6  
ID Q8CAB6 PRELIMINARY; PRT; 587 AA.  
AC Q8CAB6;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Mitogen-activated protein kinase 8 interacting protein 3  
DE (Fragment).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Hypothalamus;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
60,770 full-length cDNAs."  
RL Nature 420:563-573 (2002).  
DR EMBL; AK039130; BAC30247.1; -.  
FT NON TER 1  
SQ SEQUENCE 587 AA; 62680 MW; BCA08B87D724CBD7 CRC64;

Query Match 82.6%; Score 2870; DB 11; Length 587;  
Best Local Similarity 92.0%; Pred. No. 4.9e-200;  
Matches 540; Conservative 18; Mismatches 29; Indels 0; Gaps 0;  
QY 65 SAHTSPEKKKAKELPEMDATSSRVWILTSTLTTSKVVIIDANQPVTVDQFTVCNAHVLC 124  
Db 1 STHPSPEKKKAKETPEADATSSRVWILTSTLTTSKVVIIDANQPVTVDQFTVCNAHVLC 60  
QY 125 ISSIPAASDSYPPGEMFLDSVNPEDPGADGVLAGITLVGCATRCNVPRSNCSRRGDTTP 184

```
Db 61 ISSIPAAASDSDYPPGEMFLSDVNPEDSGADGVLGITLVGCATRCNVPRNSCSSRGDTP 120
QY 185 VLDKGQGEVATANGKVNPSQSTEEATEATEVPDPPGPSEPETATLRPGPLTEHVFTDPAP 244
Db 121 VLDKGQGDVATTANGKVNPSQSTEEATEATEVPDPPGPSESEATVTRPGPLTEHVFTDPAP 180
QY 245 TPSSGPQPGSENGPEPDSSSTRPEPEPSGDPGTGAGSSAAPTMLWGAQNGWLVHSAVANW 304
Db 181 TPSSSTQPAENGSENGTIVQVQPEPSGELSTTTSSAAPTMLWGAQNGWLVHSAVANW 240
QY 305 KKCLHSIKLSDSVSLVHVGRVVALADGTLAIFHRGEDGOWDLSNYHLMDLGHPHHSI 364
Db 241 KKCLHSIKLSDSVSLVHVGRVVALADGTLAIFHRGEDGOWDLSNYHLMDLGHPHHSI 300
QY 365 RCMAVVYDRVWCGYKXKXHVQPKTMQIEKSFDAHPRRESQVRQLAWIGDGVVWSIRLDS 424
Db 301 RCMAVVNDRVWCGYKXKXHVQPKTMQIEKSFDAHPRRESQVRQLAWIGDGVVWSIRLDS 360
QY 425 TLRLYHAHTHQHLQDVIEPYVSKMLGTGKLGFSFVRITALLVAGSRLWGTGNGVVISI 484
Db 361 TLRLYHAHTHQHLQDVIEPYVSKMLGTGKLGFSFVRITALLIAGNRLWGTGNGVVISI 420
QY 485 PLTETVVLHGRQGLLGLRANKTSPTSGEGARPGGIIHVYGDSSDRAASSFIPYCSMAQAQ 544
Db 421 PLTETVVLHGRQGLLGLRANKTSPTSGEGTRPGGIIHVYGDSSDRAASSFIPYCSMAQAQ 480
QY 545 LCFHGHRAVKKFFVSVPGNVLATLNGSVLDSPAEGPGPAAPASEVEGQKLRNVLVLSGGE 604
Db 481 LCFHGHRAVKKFFVSVPGNVLATLNGSVLDSPAEGPGPAAPADAEGQKLRNVLVLSGGE 540
QY 605 GYIDFRIGDGEDDETEEGAGDMSQVKPVLKSAERSHIIIVQVSYTPE 651
Db 541 GYIDFRIGDGEDDETEECAGDVNQTKPSLSKAERSHIIIVQVSYTPE 587

RESULT 4
Q8C7W0 PRELIMINARY; PRT; 827 AA.
AC Q8C7W0;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Similar to sperm surface protein.
GN SPAG9.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=22354683; PubMed=12466851;
RA the FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK049161; BAC33577.1; -.
DR MGD; MGI:1918084; Spag9.
SQ SEQUENCE 827 AA; 90652 MW; 06627157DDC785A0 CRC64;
```

```
Query Match 54.5%; Score 1892; DB 11; Length 827;
Best Local Similarity 57.8%; Pred. No. 8.7e-129;
Matches 391; Conservative 96; Mismatches 147; Indels 42; Gaps 17;
```

```
QY 1 MKNVPVPVYCRPLVEKDPMTKMLWCAAGVNLSGWRPNEDDAGNVKVPAP-GRDPLTCDREG 59
Db 159 MKNLPVPVYLRPLDEKASMLWCAVGNLSGGKTR--DGGSVVGASVYFKDIAGLDTEG 216
QY 60 DGEPKSAHTSPE-----KKAKELPEMDATSSRVVLTSTLTTSKVVIIDANQGVTVVD 113
Db 217 SKQRSASQSSLDKLDQELKEQKEFKNQELSSQVWICTSTHSTTKVVIIDAVQGNILD 276
```

```
QY 114 QFTVCNAHVLCISSIPAAASDSDYPPGEMFLSD-----VNPEDPGADGVLAGIT 162
Db 277 SFTVCNSHVLCIASVPGARETDYPAGEELSESGQVDKASLCGSMTSNSSAEMDSLGGIT 336
QY 163 LVGCATR-CNVPRNSCNRSGDTPVLDKGQGEVATANGKVNPSQST-EEATEATEVPDPG 220
Db 337 VVGCSTEGLTGAATSPSTNGASPVIEK-PEMET-ENSEVDENIPTAEATEATE-GNAG 393
QY 221 PSEPETATLRPGPLTEHVFTDPAPT-----SSGPQPGSENGPEPDSSSTRP-EPEPSGD 274
Db 394 STEDTVDISQPGVYTHEVFTDPLGVQIPEDLSPVFQSSNDSVDYKQISVLEPNEQDLARE 453
QY 275 PTGAGSSAAPTMLWGAQNGWLVHSAVANWKKCLHSIKLSDSVSLVHVGRVVALADG 334
Db 454 EAQKMSSLLPTMWLGAQNGCLYVHSSVAQWRKCLHSIKLSDSVSLVHVGRVVALADG 513
QY 335 TLAIHFRGEDGOWDLSNYHLMDLGHPHHSIRCMVVYDRVWCGYKXKXHVQPKTMQIEK 394
Db 514 TLAIHFRGVGDGOWDLSNYHLLDLGRPHHSIRCMTVVHDKVWCGYRNKIYVVPKAMKIEK 573
QY 395 SFDAHPRESQVRQLAWIGDGVVWSIRLDSLRLYHAHTHQHLQDVIEPYVSKMLGTGK 454
Db 574 SFDAHPRESQVRQLAWIGDGVVWSIRLDSLRLYHAHTYQHLQDVIEPYVSKMLGTGK 633
QY 455 LGFSFVRITALLVAGSRLWGTGNGVVISIPLTETVVLHGRQGLLGLRANKTSPTSGEGAR 514
Db 634 LGFSFVRITALLMVSCNRLWGTGNGVVISIPLTETVILHGRLLGLRANKTSPTGPN--R 691
QY 515 PGGIHVYGDSSDRAA--SSFIPYCSMAQAQLCFHGHRAVKKFFVSVPGNVL---ATLNG 570
Db 692 PGSVIRVYGDENSCKVTPGTFTIPYCSMAHAQQLCFHGHRAVKKFFVAVPGQVISPQSSSG 751
QY 571 SVLDSPAEGPGPAAPASEVEGQKLRNVLVLSGGEYIDFRIGDGEDDETEEGAGDMSQVK 630
Db 752 AGLTADKAGSSAQEPSSQT---PLKSMVLVISGGEYIDFRMGD-EGGESEL-LGEDLPLE 806
QY 631 PVLKSAERSHIIIVQV 646
Db 807 PSVTKAERSHLIVQV 822
```

RESULT 5

```
Q8CJC2 PRELIMINARY; PRT; 1307 AA.
AC Q8CJC2;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE JNK-associated leucine-zipper protein.
GN SPAG9.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22295031; PubMed=12391307;
RA Lee C.M., Onesime D., Reddy C.D., Dhanasekaran N., Reddy E.P.;
RT "JLP: A scaffolding protein that tethers JNK/p38MAPK signaling modules
RT and transcription factors.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14189-14194(2002).
DR EMBL; AF327451; AAN61564.1; -.
DR MGD; MGI:1918084; Spag9.
SQ SEQUENCE 1307 AA; 144592 MW; E56C26AA9DD8278C CRC64;
```

```
Query Match 52.4%; Score 1821.5; DB 11; Length 1307;
Best Local Similarity 56.4%; Pred. No. 2.1e-123;
Matches 381; Conservative 93; Mismatches 147; Indels 55; Gaps 18;
```

```
QY 1 MKNVPVPVYCRPLVEKDPMTKMLWCAAGVNLSGWRPNEDDAGNVKVPAP-GRDPLTCDREG 59
Db 652 MKNLPVPVYLRPLDEKASMLWCAVGNLSGGKTR--DGGSVVGASVYFKDIAGLDTEG 709
```



|   |   |  |
|---|---|--|
| RL  | Proc. Natl. Acad. Sci. U.S.A. 99:14189-14194 (2002).                  |  |
| DR  | EMBL; AF327452; AAN61565.1; -.  |  |
| DR  | Genew; HGNC:14524; SPAG9.   |  |
| SQ  | SEQUENCE 1311 AA; 145135 MW; D60A544D70BD1F97 CRC64;                  |  |
| Query Match   | 52.2%; Score 1814.5; DB 4; Length 1311;                               |  |
| Best Local Similarity   | 56.1%; Pred. No. 6.9e-123;  |  |
| Matches 382; Conservative 87; Mismatches 157; Indels 55; Gaps 17; |   |  |
| QY  | 1 MKNVPVPVYCRPLVEKPTMKLWCAAGVNLGWRPNEDDAGNGVKPAP-GRDPLTCDREG 59       |  |
| Db  | 656 MKNLPVPVYLRPLDEKDTSMKLWCAVGNLSGGKTR--DGGSVVGASVFYKDVAGLDTEG 713   |  |
| QY  | 60 DGEPKSAHTSPE-----KKKAKELPEMDATSSRVWILTSTLTTSKVIIDANQPGTVVD 113     |  |
| Db  | 714 SKQRSASQSSLDKLDQELKEQKQKELKNQELSSLVWICTSTHSATKVLIIIDAVQPGNILD 773 |  |
| QY  | 114 QFTVCNAHVLICISSIPAASDSYPPGEMFLDS-----VNPEDPGADGVLAGIT 162         |  |
| Db  | 774 SFTVCNSHVLCIASVPGARETDYPAGEDLSSEGVQDKASLCGSMTSNSSAETDSLGGIT 833   |  |
| QY  | 163 LVGC-ATRCNVPRSNCSRGDTPVLDKGQGEVATIANGKNVPSQST-EEATEATEVPDPG 220   |  |
| Db  | 834 VVGCSAEGVTGAATSPSTNGASPVMDKPPMEA--ENSEVDENVPTAEATEATE-GNAG 890    |  |
| QY  | 221 PSEPETATLRPGPLTEHVFTDPAPT-----SSGPQGSSENGPEPDSSSTRP-EPEPSGD 274   |  |
| Db  | 891 SAEDTVDISQTVGYTEHVFTDPLGVQIPEDLSPVYQSSNDSDAYKDQISVLPEQDLVRE 950   |  |
| QY  | 275 PTGAGSSAAPTMTWLGQNGWLYVHSAVANWKKLHSLKDSVLSLVHVKGRVLVALADG 334     |  |
| Db  | 951 EAQKMSSLLPTMTWLGQNGCLYVHSSVAQWRKCLHSLKDSILSVHVKGIVLVALADG 1010    |  |
| QY  | 335 TLAIHFRGEDQWDLNHYHLMDLGHPHHSIRCMVAVYDRVWCGYKKNKHVVIQPKTMQIEK 394  |  |
| Db  | 1011 TLAIHFRGVGDQWDLNHYHLLDLGRPHHSIRCMVAVHDKVWCGYRNKIYVVPKAMKIEK 1070 |  |
| QY  | 395 SFDAHPRESQVRQLAWIGDGVWVSIPLTETVTLHRLGQLLGRANKTSPTSSEGAR 514       |  |
| Db  | 1131 LGFSFVRITALMVSCNRLWVGTVNGVVISIPLTET-----NKTSGVPGN--R 1175        |  |
| QY  | 515 PGFIHVGDDSSDRAA--SSFIPYCSMAQAQLCFHGRDAVKFFVSPVGNVL---ATLNG 570    |  |
| Db  | 1176 PGSVIRVYDENSCKVTPCTFIPYCSMAQAQLCFHGRDAVKFFVAVPGVQVISPQSSSG 1235  |  |
| QY  | 571 SVLDSPAEGPGPAAPASEVEGQKLRNVLVLSGGEYIDFRIGDGEDDETEEGAGDMSQVK 630   |  |
| Db  | 1236 TDLTGDKAGPSAQEPGSGT---PLKSMVLVSGGEYIDFRMGD-EGGESEL-LGEDLPLE 1290 |  |
| QY  | 631 PVLSKAERSHIIIVQVSYTPE 651   |  |
| Db  | 1291 PSVTKAERSHLIVQVMYGNE 1311  |  |
| RESULT 8  |   |  |
| Q9H811  |   |  |
| ID  | Q9H811 PRELIMINARY; PRT; 734 AA.                                      |  |
| AC  | Q9H811;   |  |
| DT  | 01-MAR-2001 (TrEMBLrel. 16, Created)                                  |  |
| DT  | 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)                     |  |
| DT  | 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)                   |  |
| DE  | Hypothetical protein FLJ14006.  |  |
| OS  | Homo sapiens (Human).   |  |
| OC  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;     |  |
| OC  | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.            |  |
| OX  | NCBI_TaxID=9606;  |  |
| RN  | [1]   |  |
| RP  | SEQUENCE FROM N.A.  |  |
| RA  | Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,     |  |
| Query Match   | 18.0%; Score 623.5; DB 4; Length 200;                                 |  |
| Best Local Similarity   | 61.8%; Pred. No. 1.8e-37;   |  |
| RA  | Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,             |  |
| RA  | Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M., |  |
| RA  | Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S., |  |
| RA  | Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,    |  |
| RA  | Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,       |  |
| RA  | Ninomiya K., Iwayanagi T.,  |  |
| RT  | "NEDO human cDNA sequencing project.";                                |  |
| RL  | Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.               |  |
| DR  | EMBL; AK024068; BAB14812.1; -.  |  |
| KW  | Hypothetical protein.   |  |
| SQ  | SEQUENCE 734 AA; 79243 MW; 58270E2B3E6DAB9B CRC64;                    |  |
| Query Match   | 27.2%; Score 943; DB 4; Length 734;                                   |  |
| Best Local Similarity   | 46.8%; Pred. No. 6.5e-60;   |  |
| Matches 210; Conservative 58; Mismatches 127; Indels 54; Gaps 11; |   |  |
| QY  | 1 MKNVPVPVYCRPLVEKPTMKLWCAAGVNLGWRPNEDDAGNGVKPAP-GRDPLTCDREG 59       |  |
| Db  | 302 MKNLPVPVYLRPLDEKDTSMKLWCAVGNLSGGKTR--DGGSVVGASVFYKDVAGLDTEG 359   |  |
| QY  | 60 DGEPKSAHTSPE-----KKKAKELPEMDATSSRVWILTSTLTTSKVIIDANQPGTVVD 113     |  |
| Db  | 360 SKQRSASQSSLDKLDQELKEQKQKELKNQELSSLVWICTSTHSATKVLIIIDAVQPGNILD 419 |  |
| QY  | 114 QFTVCNAHVLICISSIPAASDSYPPGEMFLDS-----VNPEDPGADGVLAGIT 162         |  |
| Db  | 420 SFTVCNSHVLCIASVPGARETDYPAGEDLSSEGVQDKASLCGSMTSNSSAETDSLGGIT 479   |  |
| QY  | 163 LVGC-ATRCNVPRSNCSRGDTPVLDKGQGEVATIANGKNVPSQST-EEATEATEVPDPG 220   |  |
| Db  | 480 VVGCSAEGVTGAATSPSTNGASPVMDKPPMEA--ENSEVDENVPTAEATEATE-GNAG 536    |  |
| QY  | 221 PSEPETATLRPGPLTEHVFTDPAPT-----SSGPQGSSENGPEPDSSSTRP-EPEPSGD 274   |  |
| Db  | 537 SAEDTVDISQTVGYTEHVFTDPLGVQIPEDLSPVYQSSNDSDAYKDQISVLPEQDLVRE 596   |  |
| QY  | 275 PTGAGSSAAPTMTWLGQNGWLYVHSAVANWKKLHSLKDSVLSLVHVKGRVLVALADG 334     |  |
| Db  | 597 EAQKMSSLLPTMTWLGQNGCLYVHSSVAQWRKCLHSLKDSILSVHVKGIVLVALADG 656     |  |
| QY  | 335 TLAIHFRGEDQWDLNHYHLMDLGHPHHSIRCMVAVYDRVWCGYKKNKHVVIQPKTMQIEK 394  |  |
| Db  | 657 TLAIHFRGVGDQWDLNHYHLLDLGRPHHSIRCMVAVHDKVWCGYRNKIYVVHSS----- 710   |  |
| QY  | 395 SFDAHPRESQVRQLAWIGDGVWVSIPLD 423                                  |  |
| Db  | 711 -----LGDGARPCCLKTD 722  |  |
| RESULT 9  |   |  |
| Q7Z5Q8  |   |  |
| ID  | Q7Z5Q8 PRELIMINARY; PRT; 200 AA.                                      |  |
| AC  | Q7Z5Q8;   |  |
| DT  | 01-OCT-2003 (TrEMBLrel. 25, Created)                                  |  |
| DT  | 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)                     |  |
| DT  | 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)                   |  |
| DE  | Lung cancer oncogene 4.   |  |
| GN  | HLC4.   |  |
| OS  | Homo sapiens (Human).   |  |
| OC  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;     |  |
| OC  | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.            |  |
| OX  | NCBI_TaxID=9606;  |  |
| RN  | [1]   |  |
| RP  | SEQUENCE FROM N.A.  |  |
| RC  | TISSUE=Lung;  |  |
| RA  | Kim J.W.;   |  |
| RT  | "Identification of a new oncogene in human lung cancer.";             |  |
| RL  | Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.               |  |
| DR  | EMBL; AY117689; AAN12270.1; -.  |  |
| SQ  | SEQUENCE 200 AA; 21261 MW; 2F44FE28CB46174C CRC64;                    |  |







Db 673 QQLRDYRKISISFQSASASATATPEKKKLT-----PTPTPVAGGEPADAAATTGGSDTQ 726  
QY 258 PEPDSSSTRPEPE-----PSGDPT-----GAGSSAAPTMMWL 288  
Db 727 LELNLSDDDETEAAAASLVAGSTGLVERVPSPAPSYHGHQDSNPEEGESNQSTMWI 786  
QY 289 GAQNGWLYVHSAVANWKKCLHSIKLK--DSVLSLVHVKGRVLVALADGTLAIFHRGEDGQ 346  
Db 787 GTEDGCIHVYNSTDNIRIKKNRIKIEHHSAVSILYLDNRVFSLANGDLCVYLRDGATS 846  
QY 347 WDLSNYHMLDLGHPHHSIRCMAVVYDRVWCYKKNKVHVVIQPKTM-----QIEKSFDAHPRR 402  
Db 847 WNTCSSHCLSIGTVTSPVKNLLNVNGRLWCISQGIKVLVDVETLTWVNQIQISSDSKP-- 904  
QY 403 ESQVRQLAWIGDGVWVSIRLSDTLRLYHAHTHQLQDVVDIEPVVSKMLGT-----GKLGF 457  
Db 905 ---ITNMTVSNHVVWISIQNSAHIKCFHSNTHQLVTEVNLAPAVNKMLSNCDEIIRQHA 961  
QY 458 SFVRITALLVAGSRLWVGTCNGVVISIPLTETVVLHRGQLGLRANKTSPTSGEGARPGG 517  
Db 962 ACLRVTSLLCCCKDLIWIWIGTSAGVLLTIP-----AQGYEKGA 997  
QY 518 IHHVYGDDSSDRAASSFIPYCSMAQAQLCFHGHHRDAVKF--FVSVPGNVLATINGSVLDS 575  
Db 998 I-----NIVPTGIP-----HGTGHVRFLTFVETTG-----LEG 1026  
QY 576 PAEGPGPAAPASEVEG-----QKLRNVLVLSGGEGYIDFRIGDGEDDETE 620  
Db 1027 AASGAGGRDAGSSNDEYTKQGSIKHSKSKSETNNTLIISGGDGYEDF-----N 1077  
QY 621 EGAGDMSQVKPVLSKAERSHIIWQV 646  
Db 1078 SGANSLSEIAG--REDSTNHLIIWQI 1101

Search completed: August 23, 2004, 11:14:22  
Job time : 135 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2004, 00:01:48 ; Search time 1272 Seconds  
(without alignments)  
12337.141 Million cell updates/sec

Title: US-10-019-495-26  
Perfect score: 3694  
Sequence: 1 ctggcaggttcctagtgc.....gtcgtaaaaaaaaaaaaaa 3694

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002s:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description         |
|------------|--------|-------------|--------|-------|---------------------|
| 1          | 3694   | 100.0       | 3694   | 3     | AAC64282 Human mem  |
| 2          | 3496.4 | 94.7        | 4837   | 4     | AAI59280 Human pol  |
| 3          | 3494.2 | 94.6        | 4737   | 3     | AAC76926 Human ORF  |
| 4          | 3484.6 | 94.3        | 4701   | 6     | ABK09789 Human ova  |
| 5          | 3374   | 91.3        | 4956   | 4     | AAI61066 Human pol  |
| 6          | 3182.2 | 86.1        | 5007   | 5     | AAS74091 DNA encod  |
| 7          | 1531.6 | 41.5        | 4173   | 3     | AAA62899 Murine JN  |
| 8          | 1531.6 | 41.5        | 4200   | 3     | AAA62900 Murine JN  |
| 9          | 1531.6 | 41.5        | 4266   | 3     | AAA62902 Murine JN  |
| 10         | 1531.6 | 41.5        | 4269   | 3     | AAA62901 Murine JN  |
| 11         | 1457.4 | 39.5        | 2431   | 5     | AAS74089 DNA encod  |
| 12         | 486.2  | 13.2        | 4688   | 4     | ABA09028 Human sec  |
| 13         | 477    | 12.9        | 4286   | 7     | ABX63549 Human cdn  |
| 14         | 442.2  | 12.0        | 4667   | 8     | ADA45191 Murine JLP |
| 15         | 413.4  | 11.2        | 4675   | 8     | ADA45189 Human cdn  |
| 16         | 382.4  | 10.4        | 403    | 2     | AAQ39848 Expressed  |
| 17         | 382.4  | 10.4        | 403    | 2     | AAQ59260 Human bra  |
| 18         | 370.2  | 10.0        | 3113   | 2     | AAV46322 Human sec  |
| 19         | 370.2  | 10.0        | 3113   | 5     | AAF98489 Human cdn  |
| 20         | 360.6  | 9.8         | 3791   | 6     | AAS94748 Human DNA  |
| 21         | 352.4  | 9.5         | 7267   | 4     | AAK69457 Human imm  |
| 22         | 294.4  | 8.0         | 3588   | 4     | ABL13755 Drosophil  |
| 23         | 279    | 7.6         | 1966   | 4     | AAH16524 Human cdn  |

|    |       |     |       |   |          |                    |
|----|-------|-----|-------|---|----------|--------------------|
| 24 | 269.2 | 7.3 | 782   | 5 | AAS74086 | Aas74086 DNA encod |
| 25 | 269.2 | 7.3 | 1779  | 5 | AAS79794 | Aas79794 DNA encod |
| 26 | 244   | 6.6 | 633   | 4 | AAH07478 | Aah07478 Human cdn |
| 27 | 242.4 | 6.6 | 1413  | 3 | AAC98939 | Aac98939 Human pan |
| 28 | 237   | 6.4 | 237   | 3 | AAC00302 | Aac00302 Human sec |
| 29 | 226   | 6.1 | 554   | 6 | ABL90245 | Ab190245 Human pol |
| 30 | 220.4 | 6.0 | 10011 | 4 | ABL13754 | Ab113754 Drosophil |
| 31 | 219.4 | 5.9 | 675   | 6 | ABK63058 | Abk63058 Rat seque |
| 32 | 219.4 | 5.9 | 675   | 9 | ADB51737 | Adb51737 Primary r |
| 33 | 209.4 | 5.7 | 478   | 5 | AAS74070 | Aas74070 DNA encod |
| 34 | 194.8 | 5.3 | 633   | 5 | AAS84890 | Aas84890 DNA encod |
| 35 | 194.8 | 5.3 | 633   | 5 | AAS74085 | Aas74085 DNA encod |
| 36 | 194.8 | 5.3 | 633   | 5 | AAS87991 | Aas87991 DNA encod |
| 37 | 185.2 | 5.0 | 2937  | 4 | AAH18341 | Aah18341 Human cdn |
| 38 | 122.4 | 3.3 | 128   | 2 | AAT20322 | Aat20322 Human gen |
| 39 | 116.6 | 3.2 | 6289  | 6 | ABL65906 | Ab165906 Lung canc |
| 40 | 116.6 | 3.2 | 6289  | 6 | ABL65509 | Ab165509 Lung canc |
| 41 | 116.6 | 3.2 | 6289  | 6 | ABK64385 | Abk64385 Human ben |
| 42 | 116.6 | 3.2 | 7540  | 6 | ABL92090 | Ab192090 Human Tum |
| 43 | 116.6 | 3.2 | 7540  | 7 | ABX72015 | Abx72015 DNA encod |
| 44 | 116.6 | 3.2 | 7768  | 7 | ABT16473 | Abt16473 Human int |
| 45 | 108.4 | 2.9 | 731   | 5 | ABV30065 | Abv30065 Human pro |

ALIGNMENTS

RESULT 1  
AAC64282  
ID AAC64282 standard; cDNA; 3694 BP.  
XX  
AC AAC64282;  
XX  
DT 23-FEB-2001 (first entry)  
XX  
DE Human membrane-associated protein HUMAP-9 cDNA.  
XX  
KW Human membrane-associated protein; HUMAP; transgenic organism;  
KW drug screening; cell signalling modulator; agonist; antagonist;  
KW cell differentiation modulator; cell proliferation modulator;  
KW cell proliferative disorder; cancer; cell differentiation disorder;  
KW developmental disorder; cell signalling disorders; endocrine disorder;  
KW hyperpituitarism; hypothyroidism; hyperparathyroidism; infection;  
KW pancreatic disorder; diabetes mellitus; immunological disorder;  
KW hereditary neuropathy; gonadal steroid hormone associated disorder;  
KW infertility; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200065054-A2.  
XX  
PD 02-NOV-2000.  
XX  
PF 20-APR-2000; 2000WO-US010884.  
XX  
PR 23-APR-1999; 99US-0130694P.  
PR 23-JUN-1999; 99US-0140580P.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
PI Hillman JL, Bandman O, Tang YT, Lal P, Yue H, Reddy R;  
PI Azimzai Y, Baughn MR;  
XX  
DR WPI; 2000-687346/67.  
DR P-PSDB; AAB29652.  
XX  
PT Human membrane-associated protein, useful for diagnosis and treatment of  
PT cell signaling, cell differentiation and cell proliferation disorders  
PT such as cancer, and for identifying agonists and antagonists.  
XX  
PS Claim 4; Page 93; 99pp; English.  
XX  
CC The invention relates to 17 human membrane-associated proteins, HUMAP-1

CC to HUMAP-17 (AAB29644-B29660) and the cDNAs encoding them (AAC64274-  
CC C64290). The invention also relates to expression constructs, host cells  
CC and transgenic organisms comprising a HUMAP nucleic acid sequence; the  
CC recombinant preparation of a HUMAP; methods of screening compounds for  
CC their ability to modulate HUMAP activity or expression; and  
CC pharmaceutical compositions comprising a HUMAP protein, a HUMAP agonist  
CC or HUMAP antagonist. The HUMAPs acts as modulators of cell signalling,  
CC differentiation and proliferation. A HUMAP is useful for screening a  
CC compound for effectiveness as an agonist or antagonist of HUMAP activity.  
CC The protein, or the identified agonist or antagonist is useful for  
CC treating a disease or condition associated with decreased or increased  
CC expression of functional HUMAP. A HUMAP nucleic acid is useful for  
CC screening a compound for its ability to alter expression of that  
CC particular HUMAP gene. A wide variety of diseases may be treated using  
CC compositions of the invention. These diseases include cell proliferative  
CC disorders (e.g., actinic keratosis, arteriosclerosis); cancer (e.g.,  
CC breast, bladder, bone marrow, brain and uterine cancer); cell  
CC differentiation disorders, in particular developmental disorders (e.g.,  
CC renal tubular acidosis, anaemia, Cushing's syndrome, achondroplasia,  
CC epilepsy, and muscular dystrophy); cell signalling disorders, in  
CC particular endocrine disorders such as hypothalamus and pituitary  
CC disorders resulting from lesions such as thrombosis; disorders associated  
CC with hyperpituitarism (e.g., acromegaly); disorders associated with  
CC hypothyroidism (e.g., goitre); hyperparathyroidism; pancreatic disorders  
CC such as type I or type II diabetes mellitus; infections; immunological  
CC disorders; hereditary neuropathies (e.g., neurofibromatosis); and  
CC disorders associated with gonadal steroid hormones (e.g., infertility,  
CC endometriosis, polycystic ovary syndrome, osteoporosis, Leydig cell  
CC deficiency and gynecomastia). Antibodies which specifically bind HUMAP  
CC may be used for the diagnosis of disorders associated with the expression  
CC of HUMAP, or in assays to monitor patients being treated with HUMAP or  
CC agonists, antagonists or inhibitors of HUMAP. The present sequence  
CC represents a HUMAP cDNA of the invention  
XX  
SQ Sequence 3694 BP; 677 A; 1265 C; 1169 G; 583 T; 0 U; 0 Other;

Query Match 100.0%; Score 3694; DB 3; Length 3694;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3694; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGGCAGGTTCTCTAGTACGCTCTCTGGGACGGGTTGGAGGTCGAGAGGAGGCCCCG 60  
DB 1 CTGGCAGGTTCTCTAGTACGCTCTCTGGGACGGGTTGGAGGTCGAGAGGAGGCCCCG 60

QY 61 CGGCTCCACGCCCCACCTGCCTCAGTCTTGAAGTGGCTCTGCCGCCAGAGCTGGCA 120  
DB 61 CGGCTCCACGCCCCACCTGCCTCAGTCTTGAAGTGGCTCTGCCGCCAGAGCTGGCA 120

QY 121 GCCCTACCTGTCTCAGATTCTGGAGGATGGGTAGGAGCCAGGGTTCTGCCCCACGGCG 180  
DB 121 GCCCTACCTGTCTCAGATTCTGGAGGATGGGTAGGAGCCAGGGTTCTGCCCCACGGCG 180

QY 181 CCTCCCTGCTCCCTGCAGCTGAGTCCCAACGGGGGCCAGGAGGACACGGGATGAAGAAC 240  
DB 181 CCTCCCTGCTCCCTGCAGCTGAGTCCCAACGGGGGCCAGGAGGACACGGGATGAAGAAC 240

QY 241 GTGCCGGTCCGGTGTACTGCCGCCCTCTGGTGAGAGGAGCCACCATGAAGCTGTGG 300  
DB 241 GTGCCGGTCCGGTGTACTGCCGCCCTCTGGTGAGAGGAGCCACCATGAAGCTGTGG 300

QY 301 TGTGCCGGGGCGTCAACCTGAGCGGGTGGAGGGCCCAATGAGGACGAGCTGGGAATGGA 360  
DB 301 TGTGCCGGGGCGTCAACCTGAGCGGGTGGAGGGCCCAATGAGGACGAGCTGGGAATGGA 360

QY 361 GTCAGCCAGGCGCCAGGCGCGATCCCCTGACCTGCCAGCCGGAAGGAGACGGCGAGCCC 420  
DB 361 GTCAGCCAGGCGCCAGGCGCGATCCCCTGACCTGCCAGCCGGAAGGAGACGGCGAGCCC 420

QY 421 AAGAGCGCCACACGCTCTCCCGAGAAGAGAGGCCAAGGAGCTCCCTGAATGGACGCC 480  
DB 421 AAGAGCGCCACACGCTCTCCCGAGAAGAGAGGCCAAGGAGCTCCCTGAATGGACGCC 480

QY 481 ACCTCCAGCCGGGTGTGGATCCTGTACCCAGCACCCCTGACCCAGCAAGGTGGTGATCATC 540

481 ACCTCCAGCCGGGTGTGGATCCTGTACCCAGCACCCCTGACCCAGCAAGGTGGTGATCATC 540  
QY 541 GACGCCAACACAGCCGGGACGGTGGTGGACCAAGTTACCGTCTGCAACGGCACGTGCTG 600  
DB 541 GACGCCAACACAGCCGGGACGGTGGTGGACCAAGTTACCGTCTGCAACGGCACGTGCTG 600

QY 601 TGCATCTCCAGCATCCCCGGGGCCAGCGACAGCGGACTACCTCCCGGGGAGATGTTCCCTG 660  
DB 601 TGCATCTCCAGCATCCCCGGGGCCAGCGACAGCGGACTACCTCCCGGGGAGATGTTCCCTG 660

QY 661 GACAGCGACCTGAACCCAGAGAACCCCGGGCCAGATGGCGTGGTGGCCGGTATCACCCCTG 720  
DB 661 GACAGCGACCTGAACCCAGAGAACCCCGGGCCAGATGGCGTGGTGGCCGGTATCACCCCTG 720

QY 721 GTGGGCTGTGCCACCCCGCTGCAACGTCGCCGGGAGCAACTGCTCCTCCCGAGGGGACACC 780  
DB 721 GTGGGCTGTGCCACCCCGCTGCAACGTCGCCGGGAGCAACTGCTCCTCCCGAGGGGACACC 780

QY 781 CCAGTGTCTAGACAAAGGGGACAGGGGAGGTGGCCACCATCGCCAAACGGGAAGTCAACCCG 840  
DB 781 CCAGTGTCTAGACAAAGGGGACAGGGGAGGTGGCCACCATCGCCAAACGGGAAGTCAACCCG 840

QY 841 TCCAGTCCACAGAGGAGGCCACAGAGGCCACAGAGGTGCCAGACCCCTGGGCCAGCGAG 900  
DB 841 TCCAGTCCACAGAGGAGGCCACAGAGGCCACAGAGGTGCCAGACCCCTGGGCCAGCGAG 900

QY 901 CCAGAGACAGCCACATTGCGGCCCGGGCCCTCTCACAGAGCACGTTCTTCACTGACCCAGCC 960  
DB 901 CCAGAGACAGCCACATTGCGGCCCGGGCCCTCTCACAGAGCACGTTCTTCACTGACCCAGCC 960

QY 961 CCGACCCCGTCTCTGGCCCCCAGCCCTGGCAGCGAGAACGGGCCAGAGCCTGACAGCAGC 1020  
DB 961 CCGACCCCGTCTCTGGCCCCCAGCCCTGGCAGCGAGAACGGGCCAGAGCCTGACAGCAGC 1020

QY 1021 AGCACACGGCCAGAGCCAGAGCCAGCGGGGACCCACGGGAGCAGGAGCAGTGTGTCA 1080  
DB 1021 AGCACACGGCCAGAGCCAGAGCCAGCGGGGACCCACGGGAGCAGGAGCAGTGTGTCA 1080

QY 1081 CCACACCATGTGGCTGGAGCCCAAGAACGGCTGGCTCTATGTGCACTCGGCTGTGGCAAC 1140  
DB 1081 CCACACCATGTGGCTGGAGCCCAAGAACGGCTGGCTCTATGTGCACTCGGCTGTGGCAAC 1140

QY 1141 TGGAGAAGTGCCTGCATCCATCAAGCTGAAGATTCTGTGTGAGCCTGGTGTGATGTC 1200  
DB 1141 TGGAGAAGTGCCTGCATCCATCAAGCTGAAGATTCTGTGTGAGCCTGGTGTGATGTC 1200

QY 1201 AAGGCCGTGTGTGGTGGCTCTGGCGGACGGGACCCCTGGCCATCTTCCACCGTGGTGA 1260  
DB 1201 AAGGCCGTGTGTGGTGGCTCTGGCGGACGGGACCCCTGGCCATCTTCCACCGTGGTGA 1260

QY 1261 GATGCCAGTGGGATCTGAGCAACTATCACCTAATGAGACCTGGGCCACCCGCACTCC 1320  
DB 1261 GATGCCAGTGGGATCTGAGCAACTATCACCTAATGAGACCTGGGCCACCCGCACTCC 1320

QY 1321 ATCCGCTGCATGGCTGTGTGTACGACCGCGTGTGGTGGCTACAGAACCAAGGTGCAC 1380  
DB 1321 ATCCGCTGCATGGCTGTGTGTACGACCGCGTGTGGTGGCTACAGAACCAAGGTGCAC 1380

QY 1381 GTCATCCAGCCCAAGACCATGCAGATAGAGAAGTCAATTGACGCCACCCGCGGGGAG 1440  
DB 1381 GTCATCCAGCCCAAGACCATGCAGATAGAGAAGTCAATTGACGCCACCCGCGGGGAG 1440

QY 1441 AGCCAGGTGGCGAGTGGCGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCTGGAC 1500  
DB 1441 AGCCAGGTGGCGAGTGGCGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCTGGAC 1500

QY 1501 TCCACCCCTGAGGCTCTACCATGCACACACGCAACAGCATCTACAGGAGCTGGACATTGAG 1560  
DB 1501 TCCACCCCTGAGGCTCTACCATGCACACACGCAACAGCATCTACAGGAGCTGGACATTGAG 1560

QY 1561 CCCTACCTCAGCAAGATGCTAGGCACTGGCAAGCTGGGTTCTCTTGTACGATCAGC 1620

Db 1561 CCCTACGTGACGACGATGCTAGGCACTGGCAAGCTGGGTTTCTCCTTCGTACGCATCAGG 1620

Qy 1621 GCGCTGCTTGTGCGGGCAGCCGSGCTCTGGGTGGCACCGGCAACGGAGTGGTCACTCTCC 1680

Db 1621 GCGCTGCTTGTGCGGGCAGCCGSGCTCTGGGTGGCACCGGCAACGGAGTGGTCACTCTCC 1680

Qy 1681 ATCCCCCTGACAGAGACTGTGGTCTTGACACCGAGGCCAGCTCCTGGGGCTCCGAGCCAAT 1740

Db 1681 ATCCCCCTGACAGAGACTGTGGTCTTGACACCGAGGCCAGCTCCTGGGGCTCCGAGCCAAT 1740

Qy 1741 AAGACATCCCCCACCCTCTGGGAGGGCGCCCCGTCCCCGGGGGCATCCACGTGTATGGC 1800

Db 1741 AAGACATCCCCCACCCTCTGGGAGGGCGCCCCGTCCCCGGGGGCATCCACGTGTATGGC 1800

Qy 1801 GATGACAGCAGTGACAGGGCGGCCAGCAGCTTCATCCCCCTACTGCTCCATGGCCACGGCC 1860

Db 1801 GATGACAGCAGTGACAGGGCGGCCAGCAGCTTCATCCCCCTACTGCTCCATGGCCACGGCC 1860

Qy 1861 CAGCTATGCTTCCATGGGCACCGCGATGCCGTGAAGTTCTTTGTCTCGGTGCCAGGGAAC 1920

Db 1861 CAGCTATGCTTCCATGGGCACCGCGATGCCGTGAAGTTCTTTGTCTCGGTGCCAGGGAAC 1920

Qy 1921 GTGCTGGCCACCCTGAATGGCAGTGTGCTGGACAGCCCCAGCCGAGGGCCCTGGGCCAGCT 1980

Db 1921 GTGCTGGCCACCCTGAATGGCAGTGTGCTGGACAGCCCCAGCCGAGGGCCCTGGGCCAGCT 1980

Qy 1981 GCCCTGCCCTCGGAGGTCGAGGGCCAGAAAGCTGCGGAACGTGCTGGTGTGCTGAGCGCGGG 2040

Db 1981 GCCCTGCCCTCGGAGGTCGAGGGCCAGAAAGCTGCGGAACGTGCTGGTGTGCTGAGCGCGGG 2040

Qy 2041 GAGGGCTACATCGACTTCCGCAATTGGAGACGGAGAGGACGACGAGACGGAGGGCGCA 2100

Db 2041 GAGGGCTACATCGACTTCCGCAATTGGAGACGGAGAGGACGACGAGACGGAGGGCGCA 2100

Qy 2101 GGGACATGAGCCAGGTGAACCCCGTGTGTCGAGGACGAGCGCAGTCAATCATCGTG 2160

Db 2101 GGGACATGAGCCAGGTGAACCCCGTGTGTCGAGGACGAGCGCAGTCAATCATCGTG 2160

Qy 2161 TGGCAGGTGCTCTACACCCCGAGTGAAGTGTGCTGCCCTGGCCCGACCTGTACATA 2220

Db 2161 TGGCAGGTGCTCTACACCCCGAGTGAAGTGTGCTGCCCTGGCCCGACCTGTACATA 2220

Qy 2221 GGACCCCGACCACTGACCCCGCCCGCGCGGGGTAGCCAGCCAGGCGCCGCGCC 2280

Db 2221 GGACCCCGACCACTGACCCCGCCCGCGCGGGGTAGCCAGCCAGGCGCCGCGCC 2280

Qy 2281 CCTCTTCTAACTCTCAACCTGCAGCTTTACCTGAGTGTGGCCCCCTCCAGCGGGCAGGG 2340

Db 2281 CCTCTTCTAACTCTCAACCTGCAGCTTTACCTGAGTGTGGCCCCCTCCAGCGGGCAGGG 2340

Qy 2341 AGTGGGGGATGCGGATCAGCTGGGAGGAGGGGAGGGGTGCTTCCACCCGAGGGGAA 2400

Db 2341 AGTGGGGGATGCGGATCAGCTGGGAGGAGGGGAGGGGTGCTTCCACCCGAGGGGAA 2400

Qy 2401 GATGCTCTCGGGACAGTTTCCCGGGCAGTCTGGCCAGTTCAGCCAGCCAGAGTCCCTCAA 2460

Db 2401 GATGCTCTCGGGACAGTTTCCCGGGCAGTCTGGCCAGTTCAGCCAGCCAGAGTCCCTCAA 2460

Qy 2461 GTCCAGGGCACCTTGGGCCAGCGCAGGAGGAGGGGTGCTTCCAGCCAGAGTCCCTTGG 2520

Db 2461 GTCCAGGGCACCTTGGGCCAGCGCAGGAGGAGGGGTGCTTCCAGCCAGAGTCCCTTGG 2520

Qy 2521 GCCTCCTACTCCCCAGCACCTTGGGAGGAGGGGTGCTTCCCGCGCGAGGGTGCCTG 2580

Db 2521 GCCTCCTACTCCCCAGCACCTTGGGAGGAGGGGTGCTTCCCGCGCGAGGGTGCCTG 2580

Qy 2581 CCCTGGGCCACCTCTGATGCTGTGATGGGGCCACCTGCTCCTGGGCCCTCACTCT 2640

Db 2581 CCCTGGGCCACCTCTGATGCTGTGATGGGGCCACCTGCTCCTGGGCCCTCACTCT 2640

Qy 2641 GCCTAGGGAGCTGGGCCAGGCACCTAGCCTTTGCCAGGAGGTGGGCTCAGGCTGCC 2700

Db 2641 GCCTAGGGAGCTGGGCCAGGCACCTAGCCTTTGCCAGGAGGTGGGCTCAGGCTGCC 2700

RESULT 2  
AAI59280  
ID AAI59280 standard; cDNA; 4837 BP.

Qy 2701 AGGTGCCTGCACCCAGCCGGCTTCTCTGGGGCTCCCCGTCTCAAGCCTCTATCCTG 2760

Db 2701 AGGTGCCTGCACCCAGCCGGCTTCTCTGGGGCTCCCCGTCTCAAGCCTCTATCCTG 2760

Qy 2761 TCTGTCCCAACCCAGCTGTCCCCTGCCAGGAGCTGGCATAAAAGCACGAGGCCCGGC 2820

Db 2761 TCTGTCCCAACCCAGCTGTCCCCTGCCAGGAGCTGGCATAAAAGCACGAGGCCCGGC 2820

Qy 2821 TCCCTGGGGCAGCTGCTTGAGAACAGAGACTGCTACCCCATCTGCTGGGCAGGACGC 2880

Db 2821 TCCCTGGGGCAGCTGCTTGAGAACAGAGACTGCTACCCCATCTGCTGGGCAGGACGC 2880

Qy 2881 TCTTGCCAGCCCCGTTCTGACCCGTGTCCCCCAGGCTCTGCTGGGCAGAACTCACC 2940

Db 2881 TCTTGCCAGCCCCGTTCTGACCCGTGTCCCCCAGGCTCTGCTGGGCAGAACTCACC 2940

Qy 2941 TTGAGGAGTGGGCCCTTGAGTCTCTCTCCCTCCAGAAAGGGAGGCCCTTGGCCTCTCCGGGATC 3000

Db 2941 TTGAGGAGTGGGCCCTTGAGTCTCTCTCCCTCCAGAAAGGGAGGCCCTTGGCCTCTCCGGGATC 3000

Qy 3001 GGCTGCCAGGGCAGGCCAGGCTCAGGAAGAAAGGGAGGCCCTTGGCCTCTCCGGGATC 3060

Db 3001 GGCTGCCAGGGCAGGCCAGGCTCAGGAAGAAAGGGAGGCCCTTGGCCTCTCCGGGATC 3060

Qy 3061 AGTCCTAGGACACAGGCTCAGGCTCAGGTTGATGGGGATGATGTCTCCCGGGCCTGC 3120

Db 3061 AGTCCTAGGACACAGGCTCAGGCTCAGGTTGATGGGGATGATGTCTCCCGGGCCTGC 3120

Qy 3121 CTCTGTGACGGGGCTCCACGGAGCCCAGTCTCCAGACACGCTACTAAGTGCTAGGGTTG 3180

Db 3121 CTCTGTGACGGGGCTCCACGGAGCCCAGTCTCCAGACACGCTACTAAGTGCTAGGGTTG 3180

Qy 3181 CCCGTGTGGCTGTCTCCAGGGAGCAACAGAGAGGCCCAAGAGAGGCCGCTGGGGC 3240

Db 3181 CCCGTGTGGCTGTCTCCAGGGAGCAACAGAGAGGCCCAAGAGAGGCCGCTGGGGC 3240

Qy 3241 TGAGGATGAGCCCGCCCGCAGCCGACTCCAAGCCCCGAGAGGGCAGACGCCACCTGGAC 3300

Db 3241 TGAGGATGAGCCCGCCCGCAGCCGACTCCAAGCCCCGAGAGGGCAGACGCCACCTGGAC 3300

Qy 3301 TGCTCTCCCTGCCAGTGGGGCTCTCTGGCCCTATTCTTACCTTCCAGGCCACTGCAC 3360

Db 3301 TGCTCTCCCTGCCAGTGGGGCTCTCTGGCCCTATTCTTACCTTCCAGGCCACTGCAC 3360

Qy 3361 CCTGTCTGGGAGGCCCTTATGAGGGCAGCCAGCCCCGACCCCAACCCAGAGAA 3420

Db 3361 CCTGTCTGGGAGGCCCTTATGAGGGCAGCCAGCCCCGACCCCAACCCAGAGAA 3420

Qy 3421 GCACAGATCTTGGGAGCTGCCCAACAGCCCGCTGGCCACCGAGGGCTGCAGCCGCTG 3480

Db 3421 GCACAGATCTTGGGAGCTGCCCAACAGCCCGCTGGCCACCGAGGGCTGCAGCCGCTG 3480

Qy 3481 CGCTGCCGGCTTCTCCCAACCCCTGCACTCCACTGTGATGTATGTCCCTCCCTCG 3540

Db 3481 CGCTGCCGGCTTCTCCCAACCCCTGCACTCCACTGTGATGTATGTCCCTCCCTCG 3540

Qy 3541 TCTGTTCCTCCAGGATCTCGAAGTACTCCGGGCTGAGCAGTGGGGCGCTGGGGAGGG 3600

Db 3541 TCTGTTCCTCCAGGATCTCGAAGTACTCCGGGCTGAGCAGTGGGGCGCTGGGGAGGG 3600

Qy 3601 GTGACGATCTCTCAGGCTTTGGCCCTGCAAGCAAAACCCACATATCTGTATGTA 3660

Db 3601 GTGACGATCTCTCAGGCTTTGGCCCTGCAAGCAAAACCCACATATCTGTATGTA 3660

Qy 3661 ATAAATGCTTTAACGTCGTAAAAAATAAAAAA 3694

Db 3661 ATAAATGCTTTAACGTCGTAAAAAATAAAAAA 3694





QY 3540 GTCTGTTCCCCAGGATCTCGAAGTGACTCGGGGCTGAGCAGTGGGGCGGCTGGGGGAGG 3599  
|||||  
Db 4684 GTCTGTTCCCCAGGATCTCGAAGTGACTCGGGGCTGAGCAGTGGGGCGGCTGGGGGAGG 4743  
|||||  
QY 3600 GGTGACGATTCTCCTCAGGCTTTGGCCCTGCAAGCAAAACCCACATATCTGCTGTATGT 3659  
|||||  
Db 4744 GGTGACGATTCTCCTCAGGCTTTGGCCCTGCAAGCAAAACCCACATATCTGCTGTATGT 4803  
|||||  
QY 3660 AATAAATGCTTAACGTCGTAAANAAAAA 3693  
|||||  
Db 4804 AATAAATGCTTAACGTCGTAAANAAAAA 4837  
|||||

RESULT 3  
AAC76926  
ID AAC76926 standard; cDNA; 4737 BP.

XX AAC76926;

DT 08-FEB-2001 (first entry)

DE Human ORFX ORF2481 polynucleotide sequence SEQ ID NO:4961.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive; ss.

XX Homo sapiens.

OS WO200058473-A2.

PN 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US008621.

PR 31-MAR-1999; 99US-0127607P.

PR 02-APR-1999; 99US-0127636P.

PR 05-APR-1999; 99US-0127728P.

PR 30-MAR-2000; 2000US-00540763.

XX (CURA-) CURAGEN CORP.

PI Shinkets RA, Leach M;

XX WPI; 2000-602362/57.

DR P-PSDB; AAB42717.

XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease.

XX Claim 5; Page 4128-4131; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;  
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;  
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;  
CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The  
CC sequences can be used for determining the presence of or predisposition

CC to, or preventing or treating pathological conditions associated with an  
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
CC used to treat cancers, proliferative disorders, neurodegenerative  
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to  
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive  
XX  
SQ Sequence 4737 BP; 966 A; 1574 C; 1470 G; 727 T; 0 U; 0 Other;

Query Match 94.6%; Score 3494.2; DB 3; Length 4737;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 3502; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 180 GCCTCCCTGCTCCCTGCAGCTGAGTCCCAACGGGGCCAGGAGACACGGGATGAAGAA 239  
|||||  
Db 1203 GCCGGCCAAAGTACAGCAGCTGAGTCCCAACGGGGCCAGGAGACACGGGATGAAGAA 1262  
|||||  
QY 240 CGTCCCGGTGTCGGGTGTAAGTCCCGCCCTCTGTTGGAGAAGGACCCACCATTGAAGTGTG 299  
|||||  
Db 1263 CGTCCCGGTGTCGGGTGTAAGTCCCGCCCTCTGTTGGAGAAGGACCCACCATTGAAGTGTG 1322  
|||||  
QY 300 GTGTCCCGGGCGTCAACCTGAGCGGGTGGAGGCCCAATGAGGACGACGCTGGGAATGG 359  
|||||  
Db 1323 GTGTCCCGGGCGTCAACCTGAGCGGGTGGAGGCCCAATGAGGACGACGCTGGGAATGG 1382  
|||||  
QY 360 AGTCAAGCCAGCCAGCCAGGCGCGCATCCCTGACCTCGACCCGCGAAGGAGACGGCGAGCC 419  
|||||  
Db 1383 AGTCAAGCCAGCCAGGCGCGCATCCCTGACCTCGACCCGCGAAGGAGACGGCGAGCC 1442  
|||||  
QY 420 CAAGAGCGCCCAACAGCTCTCCGAGAAGAAGAGGCCCAAGGAGCTCCCTGAAATGACGCG 479  
|||||  
Db 1443 CAAGAGCGCCCAACAGCTCTCCGAGAAGAAGAGGCCCAAGGAGCTCCCTGAAATGACGCG 1502  
|||||  
QY 480 CACCTCCAGCCGGGTGTGGATCCTGACCAGCACCTTGACCACCAAGGTTGGTGATCAT 539  
|||||  
Db 1503 CACCTCCAGCCGGGTGTGGATCCTGACCAGCACCTTGACCACCAAGGTTGGTGATCAT 1562  
|||||  
QY 540 CGACGCCAACACAGCCGGGCACGGTGGTGGACCAAGTTACCGTCTGCAACCGGCACGTGCT 599  
|||||  
Db 1563 CGACGCCAACACAGCCGGGCACGGTGGTGGACCAAGTTACCGTCTGCAACCGGCACGTGCT 1622  
|||||  
QY 600 GTGCATCTCCAGCATCCCGCGGCAGCGACAGCACTACCTCCCGGGGAGATGTTCTT 659  
|||||  
Db 1623 GTGCATCTCCAGCATCCCGCGGCAGCGACAGCACTACCTCCCGGGGAGATGTTCTT 1682  
|||||  
QY 660 GGACAGCGACGTGAACCCAGAGGACCCGGGCGCAGATGGCGTGTGCGCGGTATCACCTT 719  
|||||  
Db 1683 GGACAGCGACGTGAACCCAGAGGACCCGGGCGCAGATGGCGTGTGCGCGGTATCACCTT 1742  
|||||  
QY 720 GGTGGGCTGTCCACCCGCTGCAACGTGCGCGGAGCAACTGCTCTCCCGAGGGGACAC 779  
|||||  
Db 1743 GGTGGGCTGTCCACCCGCTGCAACGTGCGCGGAGCAACTGCTCTCCCGAGGGGACAC 1802  
|||||  
QY 780 CCCAGTGTAGACAAGGGGCGAGGGGAGGTGGCCACCATCGCCAAACGGGAAGGTCAACCC 839  
|||||  
Db 1803 CCCAGTGTAGACAAGGGGCGAGGGGAGGTGGCCACCATCGCCAAACGGGAAGGTCAACCC 1862  
|||||  
QY 840 GTCCAGTCCACAGAGGAGGCCACAGAGGCCACGGAGGTGCCAGACCTTGGGGCCGCGA 899  
|||||  
Db 1863 GTCCAGTCCACAGAGGAGGCCACAGAGGCCACGGAGGTGCCAGACCTTGGGGCCGCGA 1922  
|||||  
QY 900 GCCAGAGACAGCCACATTGCGGCCCGGGCCTCTCACAGAGCACGCTTCTACTGACCCAGC 959  
|||||  
Db 1923 GCCAGAGACAGCCACATTGCGGCCCGGGCCTCTCACAGAGCACGCTTCTACTGACCCAGC 1982  
|||||  
QY 960 CCCGACCCCGTCTCTGCGCCCGCCAGCTGGCAGCGGAGACGGGCCAGAGCCTGACAGCAG 1019  
|||||  
Db 1983 CCCGACCCCGTCTCTGCGCCCGCCAGCTGGCAGCGGAGACGGGCCAGAGCCTGACAGCAG 2042  
|||||

|    |      |  |      |
|----|------|--|------|
| QY | 1020 | CAGCACACGSCCAGAGCCAGAGCCAGCGGGGACCCCACGGGAGCAGCAGCAGTGTCTGC    | 1079 |
| Db | 2043 | CAGCACACGSCCAGAGCCAGAGCCAGCGGGGACCCCACGGGAGCAGCAGCAGTGTCTGC    | 2102 |
| QY | 1080 | ACCCACCATTGTGGCTGGGAGCCCGAGAACGGCTGGCTCTATGTGCATCTGGCTGTGGCCAA | 1139 |
| Db | 2103 | ACCCACCATTGTGGCTGGGAGCCCGAGAACGGCTGGCTCTATGTGCATCTGGCTGTGGCCAA | 2162 |
| QY | 1140 | CTGGAAGAAGTGCCTGCACCTCCATCAAGCTGAAGGATTCTGTGCTGAGCCCTGGTGCATGT | 1199 |
| Db | 2163 | CTGGAAGAAGTGCCTGCACCTCCATCAAGCTGAAGGATTCTGTGCTGAGCCCTGGTGCATGT | 2222 |
| QY | 1200 | CAAAGGCCGTGTGCTGGTGGCTCTGGCGGACGGGACCCCTGGCCATCTTCCACCGTGGTGA  | 1259 |
| Db | 2223 | CAAAGGCCGTGTGCTGGTGGCTCTGGCGGACGGGACCCCTGGCCATCTTCCACCGTGGTGA  | 2282 |
| QY | 1260 | AGATGGCCAGTGGGATCTGAGCAACTATCACCTAATGGACCTGGGCCACCCGACCACTC    | 1319 |
| Db | 2283 | AGATGGCCAGTGGGATCTGAGCAACTATCACCTAATGGACCTGGGCCACCCGACCACTC    | 2342 |
| QY | 1320 | CATCCGCTGCATGGCTGTTGTGTACGACCCGCTGTGGTGTGGCTTACAGAACAAAGTGCA   | 1379 |
| Db | 2343 | CATCCGCTGCATGGCTGTTGTGTACGACCCGCTGTGGTGTGGCTTACAGAACAAAGTGCA   | 2402 |
| QY | 1380 | CGTCATCCAGCCCAAGACCATGCAGATAGAGAAATCATTTGACGCCCCACCCGCGCGGGA   | 1439 |
| Db | 2403 | CGTCATCCAGCCCAAGACCATGCAGATAGAGAAATCATTTGACGCCCCACCCGCGCGGGA   | 2462 |
| QY | 1440 | GAGCCAGGTGCGGCAGCTGGCGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCTTGA    | 1499 |
| Db | 2463 | GAGCCAGGTGCGGCAGCTGGCGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCTTGA    | 2522 |
| QY | 1500 | CTCCACCCCTGAGGCTCTACCATGCACACACGCCACCAGCATCTACAGGACGTGGACATTGA | 1559 |
| Db | 2523 | CTCCACCCCTGAGGCTCTACCATGCACACACGCCACCAGCATCTACAGGACGTGGACATTGA | 2582 |
| QY | 1560 | GCCCTACGTGAGCAAGATGCTAGGCACCTGGCAAGCTGGTTTCTCTTCTGTCGCATCAC    | 1619 |
| Db | 2583 | GCCCTACGTGAGCAAGATGCTAGGCACCTGGCAAGCTGGTTTCTCTTCTGTCGCATCAC    | 2642 |
| QY | 1620 | GGCCCTGCTTGTGCGCGGCAGCCGGCTCTGGGTGGGCAACGGCAACGGAGTGTCTCTC     | 1679 |
| Db | 2643 | GGCCCTGCTTGTGCGCGGCAGCCGGCTCTGGGTGGGCAACGGCAACGGAGTGTCTCTC     | 2702 |
| QY | 1680 | CATCCCCCTGACAGAGACTGTGGTCCTGACCCGAGGCGGCTCTGGGGCTCCGAGCCAA     | 1739 |
| Db | 2703 | CATCCCCCTGACAGAGACTGTGGTCCTGACCCGAGGCGGCTCTGGGGCTCCGAGCCAA     | 2762 |
| QY | 1740 | TAAGACATCCCCCACTCTGGGAGGGCGGCCGCTCCCGGGGGCATCATCCAGTGTATGG     | 1799 |
| Db | 2763 | TAAGACATCCCCCACTCTGGGAGGGCGGCCGCTCCCGGGGGCATCATCCAGTGTATGG     | 2822 |
| QY | 1800 | CGATGACAGCAGTGAAGGGCGGCCAGCAGCTTCATCCCCCTACTGCTCCATGGCCCCAGGC  | 1859 |
| Db | 2823 | CGATGACAGCAGTGAAGGGCGGCCAGCAGCTTCATCCCCCTACTGCTCCATGGCCCCAGGC  | 2882 |
| QY | 1860 | CCAGCTATGCTTCCATGGGCACCGCGATCCCGTGAAAGTTCTTTGTCTCGGTGCCAGGGAA  | 1919 |
| Db | 2883 | CCAGCTATGCTTCCATGGGCACCGCGATCCCGTGAAAGTTCTTTGTCTCGGTGCCAGGGAA  | 2942 |
| QY | 1920 | CGTGTGGCCACCCCTGAATGGCAGTGTGTGACAGCCAGCCGAGGGCCCTGGGCCAGC      | 1979 |
| Db | 2943 | CGTGTGGCCACCCCTGAATGGCAGTGTGTGACAGCCAGCCGAGGGCCCTGGGCCAGC      | 3002 |
| QY | 1980 | TGCCCCCTGCCTCGGAGGTGAGGGCCAGAAGCTGCGGAAAGTGTGTGTGAGCGGCGG      | 2039 |
| Db | 3003 | TGCCCCCTGCCTCGGAGGTGAGGGCCAGAAGCTGCGGAAAGTGTGTGTGAGCGGCGG      | 3062 |
| QY | 2040 | GGAGGGCTACATCGACTTCCGATTGGAGACGGAGAGGACACGAGACGGAGGAGGCGC      | 2099 |
| Db | 3063 | GGAGGGCTACATCGACTTCCGATTGGAGACGGAGAGGACACGAGACGGAGGAGGCGC      | 3122 |

|    |      |   |      |
|----|------|---|------|
| QY | 2100 | AGGGACATGAGCCAGGTGAAGCCCGTGTCTCAAGGCAGAGCGAGTCACATCATCGT      | 2150 |
| Db | 3123 | AGGGACATGAGCCAGGTGAAGCCCGTGTCTCAAGGCAGAGCGAGTCACATCATCGT      | 3182 |
| QY | 2160 | GTGGCAGGTGTCTTACACCCCGAGTGAAGCTGTGCTGCCCTGCTGGCCCCGACCTGTACAT | 2219 |
| Db | 3183 | GTGGCAGGTGTCTTACACCCCGAGTGAAGCTGTGCTGCCCTGCTGGCCCCGACCTGTACAT | 3242 |
| QY | 2220 | AGGACCCCGACACCTGACCCCGCCCGCCGCGCGGGGTAGCCAGCCAGGCGCCGCGC      | 2279 |
| Db | 3243 | AGGACCCCGACACCTGACCCCGCCCGCCGCGCGGGGTAGCCAGCCAGGCGCCGCGC      | 3302 |
| QY | 2280 | CCCTCTTCTAAACCTCTCAACCTGCAGCTTTCACTGTAGTCTGGCCCTCCAGCGGCAGG   | 2339 |
| Db | 3303 | CCCTCTTCTAAACCTCTCAACCTGCAGCTTTCACTGTAGTCTGGCCCTCCAGCGGCAGG   | 3362 |
| QY | 2340 | GAGTCGGGGATGCGGATCAGCTGGGAGGAGGAGGGGTGCTTCCACCCGAGGGGA        | 2399 |
| Db | 3363 | GAGTCGGGGATGCGGATCAGCTGGGAGGAGGAGGGGTGCTTCCACCCGAGGGGA        | 3422 |
| QY | 2400 | AGATGTCTCGGGACAGTTTCCCGGCAGCTCTGGCCAGCTTCAGCCCCAGAGTCTCTCA    | 2459 |
| Db | 3423 | AGATGTCTCGGGACAGTTTCCCGGCAGCTCTGGCCAGCTTCAGCCCCAGAGTCTCTCA    | 3482 |
| QY | 2460 | AGTCCAGGGCACCTTGGGCCAGCGCAGGAGGAGGAGTCCGAGTGTCTTGGCTTACCCCTG  | 2519 |
| Db | 3483 | AGTCCAGGGCACCTTGGGCCAGCGCAGGAGGAGGAGTCCGAGTGTCTTGGCTTACCCCTG  | 3542 |
| QY | 2520 | GGCTCCTACTCCCCAGCACCCCTGGAGAGGCAGGGGTCCCGCCCGCAGGCTGCCT       | 2579 |
| Db | 3543 | GGCTCCTACTCCCCAGCACCCCTGGAGAGGCAGGGGTCCCGCCCGCAGGCTGCCT       | 3602 |
| QY | 2580 | GCCCTGGGCCCACTCTGCATGTCTCATTGGGGCCACCCCTGCCTCTGGGCCCTCACTC    | 2639 |
| Db | 3603 | GCCCTGGGCCCACTCTGCATGTCTCATTGGGGCCACCCCTGCCTCTGGGCCCTCACTC    | 3662 |
| QY | 2640 | TGCCTAGGGGAGCTGGGCCAGGCACCTAGCCTTGGCCAGGAGGTGGCCCTCAGGCTGCC   | 2699 |
| Db | 3663 | TGCCTAGGGGAGCTGGGCCAGGCACCTAGCCTTGGCCAGGAGGTGGCCCTCAGGCTGCC   | 3722 |
| QY | 2700 | CAGTGCCTGCACCCAGCCGGCCTTCTCTGGGGCCTCCCGTCTCAAGCCTCTATCCT      | 2759 |
| Db | 3723 | CAGTGCCTGCACCCAGCCGGCCTTCTCTGGGGCCTCCCGTCTCAAGCCTCTATCCT      | 3782 |
| QY | 2760 | GTCTGTCCCCACCCAGCTGTCCCTGCCAGGGAGCTGGCATATAAAGCACGAGGCCCGG    | 2819 |
| Db | 3783 | GTCTGTCCCCACCCAGCTGTCCCTGCCAGGGAGCTGGCATATAAAGCACGAGGCCCGG    | 3842 |
| QY | 2820 | CTCCTGGGGCAGCTGTCTTGAGAACAGAGACTGTACCCCATCTCTGCCATGCAGGCAGG   | 2879 |
| Db | 3843 | CTCCTGGGGCAGCTGTCTTGAGAACAGAGACTGTACCCCATCTCTGCCATGCAGGCAGG   | 3902 |
| QY | 2880 | CTCTTGCCAGCCCGTCTTGACCCGTCTCCCCAGGTCTGCCCTGGGCAGAACTCAC       | 2939 |
| Db | 3903 | CTCTTGCCAGCCCGTCTTGACCCGTCTCCCCAGGTCTGCCCTGGGCAGAACTCAC       | 3962 |
| QY | 2940 | CTTGGAGGAGTGGGCCCTGGAGTCTGTCTCTCCAGAGCCCCCAGGTGGGATTTCTC      | 2999 |
| Db | 3963 | CTTGGAGGAGTGGGCCCTGGAGTCTGTCTCTCCAGAGCCCCCAGGTGGGATTTCTC      | 4022 |
| QY | 3000 | AGGCTGCCAGGGCAGGCCACGGCTCAGGAAGAGGGAGGCCCTTGGCTCTCCGGGAT      | 3059 |
| Db | 4023 | AGGCTGCCAGGGCAGGCCACGGCTCAGGAAGAGGGAGGCCCTTGGCTCTCCGGGAT      | 4082 |
| QY | 3060 | CAGTCTAGGACACAGGCTCAGCTCAGGTTGATGGGGATGATGTCTCCCGGGCCTG       | 3119 |
| Db | 4083 | CAGTCTAGGACACAGGCTCAGCTCAGGTTGATGGGGATGATGTCTCCCGGGCCTG       | 4142 |
| QY | 3120 | CCTCTGCACGGGCTCCACGGAGCCAGCTCCAGACACGCTACTAAGTGCCTAGGGTT      | 3179 |
| Db | 4143 | CCTCTGCACGGGCTCCACGGAGCCAGCTCCAGACACGCTACTAAGTGCCTAGGGTT      | 4202 |
| QY | 3180 | GCCCCGTGTGGCTGTCTCCAGGAGCAACAGAGAGGCCCAACAGAGAGGCCCTGGGG      | 3239 |

```
Db 4203 GCCCGCTGTGGCTGCTCCAGGGAGCAACAGAGAGGCCACCAAGCAGAGGCCCGTGGG 4262
QY 3240 CTGAGGATGAGCGCGCCCGCCAGCCGACTCCAAGCCCGCAGAGGGCAGACGCCACCTGGA 3299
Db 4263 CTGAGGATGAGCGCGCCCGCCAGCCGACTCCAAGCCCGCAGAGGGCAGACGCCACCTGGA 4322
QY 3300 CTGCTCTCCCTGCCAGCTGGGCTCTCTGGCTATTCTTACCTTCCAGGCCACTGCAC 3359
Db 4323 CTGCTCTCCCTGCCAGCTGGGCTCTCTGGCTATTCTTACCTTCCAGGCCACTGCAC 4382
QY 3360 TCCTGTCTGGAGGCCCTTATGAGGGCAGCCAGCCCGCCGACCCACCCCAACAGAGA 3419
Db 4383 TCCTGTCTGGAGGCCCTTATGAGGGCAGCCAGCCCGCCGACCCACCCCAACAGAGA 4442
QY 3420 AGCACAGATCTTGGGAGCTGCCCAAGCCCGCTGGCCACCGAGGGCTGCAGCCGCT 3479
Db 4443 AGCACAGATCTTGGGAGCTGCCCAAGCCCGCTGGCCACCGAGGGCTGCAGCCGCT 4502
QY 3480 GCGCTGCCGGCTTCTCCCCACCACTGCTCCACTGTGATGTATGTCCGCTCCCTC 3539
Db 4503 GCGCTGCCGGCTTCTCCCCACCACTGCTCCACTGTGATGTATGTCCGCTCCCTC 4562
QY 3540 GTCTGTTCCCCAGGATCTCGAAGTGACTCCGGGCTGAGCAGTGGGGCGGCTGGGGAGG 3599
Db 4563 GTCTGTTCCCCAGGATCTCGAAGTGACTCCGGGCTGAGCAGTGGGGCGGCTGGGGAGG 4622
QY 3600 GGTGACGATTCTCCTCAGGCTTTGGCCCTGCAAGCAACCCACATATCTGCTGTATGT 3659
Db 4623 GGTGACGATTCTCCTCAGGCTTTGGCCCTGCAAGCAACCCACATATCTGCTGTATGT 4682
QY 3660 AATAAATGCTTAACGTGCTAAAAA 3694
Db 4683 AATAAATGCTTAACGTCAAAAAA 4717
```

RESULT 4

ABK09789  
ID ABK09789 standard; cDNA; 4701 BP.

XX AC ABK09789;

XX D/T 14-MAR-2002 (first entry)

XX DE Human ovarian tumour protein encoding cDNA #322.

XX KW Human; ovarian tumour protein; cancer; cytostatic; immunostimulant; ss;  
KW gene therapy; CD4+ T cell; CD8+ T cell; PCR primer.

XX OS Homo sapiens.

XX PN WO200190154-A2.

XX PD 29-NOV-2001.

XX PF 23-MAY-2001; 2001WO-US016895.

XX PR 24-MAY-2000; 2000US-0207107P.

XX PR 13-JUN-2000; 2000US-0211457P.

XX PR 21-JUN-2000; 2000US-0213673P.

XX PR 03-AUG-2000; 2000US-0223288P.

XX PR 01-MAR-2001; 2001US-0272790P.

XX PA (CORI-) CORIXA CORP.

XX PI Xu J, Mitcham JL, Harlocker SL, Dillon DC, Secrist H, Lodes MJ;

XX PI Algate PA, Fling SP, Mannion J, Benson DR, Carter D;

XX XX WPI; 2002-097641/13.

XX XX New isolated polynucleotide encoding polypeptide comprising portion of  
PT ovarian tumor protein, useful for detection, diagnosis and therapy of  
PT human ovarian cancer.

```
XX PS Claim 1; Page 266-268; 285pp; English.
XX CC The invention relates to an isolated polynucleotide encoding a
CC polypeptide comprising a portion of an ovarian tumour protein. The
CC sequences of the invention are useful for stimulating an immune response
CC and for treating ovarian cancer in a patient. An antigen presenting cell
CC that expresses the sequences is useful for treating ovarian cancer by
CC incubating CD4+ and/or CD8+ T cells isolated from a patient. The T cells
CC can then be proliferated and administered to the patient to inhibit the
CC development of cancer. The DNA sequences are useful as probes or primers
CC for nucleic acid hybridisation, to direct expression of a polypeptide in
CC appropriate host cells. Detecting the presence of a cancer in a patient
CC involves obtaining a biological sample from the patient, contacting the
CC biological sample with an agent that binds to the protein, detecting the
CC amount of protein that binds to the agent, comparing the amount of
CC protein to a predetermined cut-off value and determining the presence of
CC cancer. Sequences ABK09464-ABK09802 represent PCR primers and cDNA
CC molecules encoding ovarian tumour proteins of the invention
XX SQ Sequence 4701 BP; 928 A; 1574 C; 1471 G; 728 T; 0 U; 0 Other;

Query Match 94.3%; Score 3484.6; DB 6; Length 4701;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3490; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 180 GCCTCCCTGCTCCCTGCAGCTGAGTCCCAACGGGGGCCAGGAGCACCGGATGAAGAA 239
Db 1203 GCCCGCCAAAGTACAAGCAGCTGAGTCCCAACGGGGGCCAGGAGCACCGGATGAAGAA 1262
QY 240 CGTGCCGGTGCCGGTGTA CTGCGGCCCTCTGTTGGAGAAAGAGACCCACCATGAGCTGTG 299
Db 1263 CGTGCCGGTGCCGGTGTA CTGCGGCCCTCTGTTGGAGAAAGAGACCCACCATGAGCTGTG 1322
QY 300 GTGTGCCCGGGCGTCAA CCTGAGCGGGTGGAGGCCCAATGAGGACGACGCTGGGAATGG 359
Db 1323 GTGTGCCCGGGCGTCAA CCTGAGCGGGTGGAGGCCCAATGAGGACGACGCTGGGAATGG 1382
QY 360 AGTCAAGCCAGCGCCAGGCCGCGATCCCTGACCTGCGACCGCGAAGGAGACGCGAGCC 419
Db 1383 AGTCAAGCCAGCGCCAGGCCGCGATCCCTGACCTGCGACCGCGAAGGAGACGCGAGCC 1442
QY 420 CAAGAGCGCCACACAGCTCTCCGAGAAGAAGAGGCCAAGAGCTCCCTGAAATGGACGC 479
Db 1443 CAAGAGCGCCACACAGCTCTCCGAGAAGAAGAGGCCAAGAGCTCCCTGAAATGGACGC 1502
QY 480 CACCTCCAGCCGGGTGTGGATCCTGACCAGCACCTGTGACCAAGGTGGTGTATCAT 539
Db 1503 CACCTCCAGCCGGGTGTGGATCCTGACCAGCACCTGTGACCAAGGTGGTGTATCAT 1562
QY 540 CGACGCCAAACCAAGCCGGGCACGGTGGTGGAC CAGATTACCGTCTGCAACCGGACGTGCT 599
Db 1563 CGACGCCAAACCAAGCCGGGCACGGTGGTGGAC CAGATTACCGTCTGCAACCGGACGTGCT 1622
QY 600 GTGCATCTCCAGCATCCCGCGGCCAGCGACAGCGACTA CCGTCCCGGGGAGATGTTCTT 659
Db 1623 GTGCATCTCCAGCATCCCGCGGCCAGCGACAGCGACTA CCGTCCCGGGGAGATGTTCTT 1682
QY 660 GGACAGCGACGTGAACCCAGAGACCCCGGGCGGAGATGGCGTGTGCGCCGGTATCACCTT 719
Db 1683 GGACAGCGACGTGAACCCAGAGACCCCGGGCGGAGATGGCGTGTGCGCCGGTATCACCTT 1742
QY 720 GGTGGGCTGTGCCACCCGCTGCAACGCTGCCGCGAGCAACTGCTCTCTCCGAGGGGACAC 779
Db 1743 GGTGGGCTGTGCCACCCGCTGCAACGCTGCCGCGAGCAACTGCTCTCTCCGAGGGGACAC 1802
QY 780 CCCAGTGTAGACAAGGGGCGAGGGGAGGTGGCCACCATCGCCAAACGGGAAGGTCAACCC 839
Db 1803 CCCAGTGTAGACAAGGGGCGAGGGGAGGTGGCCACCATCGCCAAACGGGAAGGTCAACCC 1862
QY 840 GTCCAGTCCACAGAGGAGGCCACAGAGGCCACCGAGGTTGCCAGACCTTGGGCCCGAGCGA 899
Db 1863 GTCCAGTCCACAGAGGAGGCCACAGAGGCCACCGAGGTTGCCAGACCTTGGGCCCGAGCGA 1922
```

QY 900 GCCAGACAGCCACATTGCGGCCCGGCTCTCAAGAGCACGTTCTCACTGACCCAGC 959  
Db 1923 GCCAGACAGCCACATTGCGGCCCGGCTCTCAAGAGCACGTTCTCACTGACCCAGC 1982  
QY 960 CCCGACCCCGTCTCTGCGGCCCGGCTGGCAGCGAGAACGGGCCAGAGCCCTGACAGCAG 1019  
Db 1983 CCCGACCCCGTCTCTGCGGCCCGGCTGGCAGCGAGAACGGGCCAGAGCCCTGACAGCAG 2042  
QY 1020 CAGCACAGGCCAGAGCCAGAGCCAGCGGGGACCCACGGGAGCAGGACAGTGTCTGC 1079  
Db 2043 CAGCACAGGCCAGAGCCAGAGCCAGCGGGGACCCACGGGAGCAGGACAGTGTCTGC 2102  
QY 1080 ACCACCATGTGGTGGGAGCCCAAGAACGGCTGCTCTATGTGCACTCGGCTGTGGCCAA 1139  
Db 2103 ACCACCATGTGGTGGGAGCCCAAGAACGGCTGCTCTATGTGCACTCGGCTGTGGCCAA 2162  
QY 1140 CTGGAAGAAAGTGCCTGCACTCCATCAAGCTGAAGATTTCTGTGTGAGCCTGTGTGCAATGT 1199  
Db 2163 CTGGAAGAAAGTGCCTGCACTCCATCAAGCTGAAGATTTCTGTGTGAGCCTGTGTGCAATGT 2222  
QY 1200 CAAAGGCCGTGTGCTGGTGGCTCTGGCGGACGGGACCCCTGGCCATCTTCCACCGTGTGA 1259  
Db 2223 CAAAGGCCGTGTGCTGGTGGCTCTGGCGGACGGGACCCCTGGCCATCTTCCACCGTGTGA 2282  
QY 1260 AGATGGCCAGTGGGATCTGAGCAACTATCACCTAATGGACCTGGGCCACCCGACCACTC 1319  
Db 2283 AGATGGCCAGTGGGATCTGAGCAACTATCACCTAATGGACCTGGGCCACCCGACCACTC 2342  
QY 1320 CATCCGCTGCATGGCTGTTGTGTACGACCGCTGTGGTGTGGGTACAGAACAAAGGTGCA 1379  
Db 2343 CATCCGCTGCATGGCTGTTGTGTACGACCGCTGTGGTGTGGGTACAGAACAAAGGTGCA 2402  
QY 1380 CGTCATCCAGCCCAAGACCAATGCAGATAGAGAGTCAATTTGACGCCCCACCCGCGCGGGA 1439  
Db 2403 CGTCATCCAGCCCAAGACCAATGCAGATAGAGAGTCAATTTGACGCCCCACCCGCGCGGGA 2462  
QY 1440 GAGCCAGGTGCGGCAGCTGGCGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCTGGA 1499  
Db 2463 GAGCCAGGTGCGGCAGCTGGCGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCTGGA 2522  
QY 1500 CTCCACCTGAGGCTCTACCATGCACACACGACCAAGCTATACAGGACGTGGACATTGA 1559  
Db 2523 CTCCACCTGAGGCTCTACCATGCACACACGACCAAGCTATACAGGACGTGGACATTGA 2582  
QY 1560 GCCCTAGTCAGCAAGATGCTAGGCATCTGCAAGCTGGGTTCTCCTTCTGACGATCAC 1619  
Db 2583 GCCCTAGTCAGCAAGATGCTAGGCATCTGCAAGCTGGGTTCTCCTTCTGACGATCAC 2642  
QY 1620 GGCCTGTTGTGCGGGCAGCCGGCTCTGGGTGGGACCGGCAACGGAGTGGTCACTC 1679  
Db 2643 GGCCTGTTGTGCGGGCAGCCGGCTCTGGGTGGGACCGGCAACGGAGTGGTCACTC 2702  
QY 1680 CATCCCTCTGACAGAGTGTGGTCTTGACCGAGGCAAGCTCCTGGGCTCCGAGCCAA 1739  
Db 2703 CATCCCTCTGACAGAGTGTGGTCTTGACCGAGGCAAGCTCCTGGGCTCCGAGCCAA 2762  
QY 1740 TAAGACATCCCCACCTCTGGGAGGGCGCCCGTCCCGGGGCATCATCCACGTGTATGG 1799  
Db 2763 TAAGACATCCCCACCTCTGGGAGGGCGCCCGTCCCGGGGCATCATCCACGTGTATGG 2822  
QY 1800 CGATGACAGCAGTACAGGGCGGCCAGCAGCTTCACTCCCTACTGCTCCATGGCCCCAGGC 1859  
Db 2823 CGATGACAGCAGTACAGGGCGGCCAGCAGCTTCACTCCCTACTGCTCCATGGCCCCAGGC 2882  
QY 1860 CCAGCTATGCTTCCATGGGCACCGGATGCCGTGAAGTTCTTTGTCTCGGTGCCAGGAA 1919  
Db 2883 CCAGCTATGCTTCCATGGGCACCGGATGCCGTGAAGTTCTTTGTCTCGGTGCCAGGAA 2942  
QY 1920 CGTGTGGCCACCCCTGAATGGCAGTGTCTGGACAGCCCGAGGCGCCCTGCGGCAGC 1979  
Db 2943 CGTGTGGCCACCCCTGAATGGCAGTGTCTGGACAGCCCGAGGCGCCCTGCGGCAGC 3002

QY 1980 TGCCCTGCCTCGAGGTCGAGGCGCAGAAAGCTGCGAAACGTGCTGTGCTGAGCGCGG 2039  
Db 3003 TGCCCTGCCTCGAGGTCGAGGCGCAGAAAGCTGCGAAACGTGCTGTGCTGAGCGCGG 3062  
QY 2040 GGAGGCTACATCGACTTCGCAATTGGAGACGAGAGGACGACGAGCGAGGAGGCGC 2099  
Db 3063 GGAGGCTACATCGACTTCGCAATTGGAGACGAGAGGACGACGAGCGAGGAGGCGC 3122  
QY 2100 AGGGACATGAGCCAGGTGAAGCCGCTGTGTCCAAGGAGAGCGCAGTCACATCATGT 2159  
Db 3123 AGGGACATGAGCCAGGTGAAGCCGCTGTGTCCAAGGAGAGCGCAGTCACATCATGT 3182  
QY 2160 GTGGCAGGTGTCTTACACCCCGAGTGAAGCTGTGCTGCCCTGCCCTGAGCCCTGTACAT 2219  
Db 3183 GTGGCAGGTGTCTTACACCCCGAGTGAAGCTGTGCTGCCCTGCCCTGAGCCCTGTACAT 3242  
QY 2220 AGGACCCCGACACCTGACCCCGCCCGCCCGGGGTAGCCAGCCAGGCGCGCCCGC 2279  
Db 3243 AGGACCCCGACACCTGACCCCGCCCGCCCGGGGTAGCCAGCCAGGCGCGCCCGC 3302  
QY 2280 CCTCTTCTAACCCTCAACCTGCAGCTTTTCACTGAGTCTGCGCCCTCCAGCGGCGAGG 2339  
Db 3303 CCTCTTCTAACCCTCAACCTGCAGCTTTTCACTGAGTCTGCGCCCTCCAGCGGCGAGG 3362  
QY 2340 GAGTGGGGATGCGGATCAGCTGGGAGGAGGAGGGGGTGTCTTCCACCCGAGGGGA 2399  
Db 3363 GAGTGGGGATGCGGATCAGCTGGGAGGAGGAGGGGGTGTCTTCCACCCGAGGGGA 3422  
QY 2400 AGATGCTCTCGGACAGTTTCCCGGGCAGCTTCCGCGCAGCTTCCAGCCAGAGTCTCTCA 2459  
Db 3423 AGATGCTCTCGGACAGTTTCCCGGGCAGCTTCCGCGCAGCTTCCAGCCAGAGTCTCTCA 3482  
QY 2460 AGTCCAGGCACTTGGGCCAGCGCAGGAGGAGGAGGAGGAGGAGTCTGGCTTACCCTG 2519  
Db 3483 AGTCCAGGCACTTGGGCCAGCGCAGGAGGAGGAGGAGGAGGAGTCTGGCTTACCCTG 3542  
QY 2520 GGCCTCTACTCCCCAGACACCCCTGGAGAGGCGAGGGCTCCCCCGCCGCGAGGCTGCCT 2579  
Db 3543 GGCCTCTACTCCCCAGACACCCCTGGAGAGGCGAGGGGTCCCCCGCCGCGAGGCTGCCT 3602  
QY 2580 GGCCTGGSCCACCCTCTGCATGCTGCTCATGGGGCCACCTGCTGCTGGGCCCTCACTC 2639  
Db 3603 GGCCTGGSCCACCCTCTGCATGCTGCTCATGGGGCCACCTGCTGCTGGGCCCTCACTC 3662  
QY 2640 TGCCTAGGGAGCTGGSCCAGGCACTAGCCTTTGCCAGGAGGTGGGCTCAGGCTGCC 2699  
Db 3663 TGCCTAGGGAGCTGGSCCAGGCACTAGCCTTTGCCAGGAGGTGGGCTCAGGCTGCC 3722  
QY 2700 CAGTGCTGTCACCCCAGCCCGGCTTCTTGGGGCTTCCCGTCTGCTCAAGCCTCTATCCT 2759  
Db 3723 CAGTGCTGTCACCCCAGCCCGGCTTCTTGGGGCTTCCCGTCTGCTCAAGCCTCTATCCT 3782  
QY 2760 GTCTGTCCCCACCCACAGCTGTCCCCCTGCCAGGAGTGGCATAAAGCACGAGGCCCGG 2819  
Db 3783 GTCTGTCCCCACCCACAGCTGTCCCCCTGCCAGGAGTGGCATAAAGCACGAGGCCCGG 3842  
QY 2820 CTCCCTGGGGCAGCTGTGAGAACAGAGACTGCTACCCCATCTGCCCCATGCGGCGAGG 2879  
Db 3843 CTCCCTGGGGCAGCTGTGAGAACAGAGACTGCTACCCCATCTGCCCCATGCGGCGAGG 3902  
QY 2880 CTCTTGGCCAGCCCGTCTGACCCGCTTCCCCCAGGCTTGGCTGGGCAGAGACTCAC 2939  
Db 3903 CTCTTGGCCAGCCCGTCTGACCCGCTTCCCCCAGGCTTGGCTGGGCAGAGACTCAC 3962  
QY 2940 CTTGAGGAGTGGGCCCTGGAGTCTGTCCCTCCAGAAAGCCCGGAGGTTGGATTCTC 2999  
Db 3963 CTTGAGGAGTGGGCCCTGGAGTCTGTCCCTCCAGAAAGCCCGGAGGTTGGATTCTC 4022  
QY 3000 AGGTGCCAGGCGGCGCCAGGCTCAGGAAGAGGAGGCGCCCTGCGCTCTCCGGAT 3059  
Db 4023 AGGTGCCAGGCGGCGCCAGGCTCAGGAAGAGGAGGCGCCCTGCGCTCTCCGGAT 4082  
QY 3060 CAGTCTAGGACAGGCTCAGCCTCAGGTTGATGGGGGATGATGTCTCCCGGGGCTG 3119

```
Db 4083 CAGTCCTAGGACACAGGCTCAGCCTCAGGTTGATGGGGATGATGTCTCCCGGGCCTG 4142
QY 3120 CCTCTGACGGGGCTCCAGGAGCCAGCTCCAGACACGCTACTAAGTGCCTAGGGTT 3179
Db 4143 CCTCTGACGGGGCTCCAGGAGCCAGCTCCAGACACGCTACTAAGTGCCTAGGGTT 4202
QY 3180 GCCCGCTGTGGCCTGCTCCAGGGAGCAACAGAGAGGCCCAAGCAGAGAGCGCCGCTGGG 3239
Db 4203 GCCCGCTGTGGCCTGCTCCAGGGAGCAACAGAGAGGCCCAAGCAGAGAGCGCCGCTGGG 4262
QY 3240 CTGAGGATGGAGCGCCGCCAGCCGACTCCAGGCCGCGAGAGGGCAGACGCCACCTGGA 3299
Db 4263 CTGAGGATGGAGCGCCGCCAGCCGACTCCAGGCCGCGAGAGGGCAGACGCCACCTGGA 4322
QY 3300 CTGCTCTCCCTGCCAGCTGGGCTCTCTGGCTATTCTTCCAGGCTCCAGGCTCCAG 3359
Db 4323 CTGCTCTCCCTGCCAGCTGGGCTCTCTGGCTATTCTTCCAGGCTCCAGGCTCCAG 4382
QY 3360 TCCTGTCTGGAGGCCCTTATGAGGGCAGCCAGCCCCCGCAGCCACCCCAACCAGAGA 3419
Db 4383 TCCTGTCTGGAGGCCCTTATGAGGGCAGCCAGCCCCCGCAGCCACCCCAACCAGAGA 4442
QY 3420 AGCACAGATCTTGGGGAGCTGCCCAACAAGCCCGCTGGCCACCGAGGGCTGCAGCCGCT 3479
Db 4443 AGCACAGATCTTGGGGAGCTGCCCAACAAGCCCGCTGGCCACCGAGGGCTGCAGCCGCT 4502
QY 3480 GCGCTGCCGCTTCTCCCAACCACTCCACTGTGATGTATGTCCGCTCCCTC 3539
Db 4503 GCGCTGCCGCTTCTCCCAACCACTCCACTGTGATGTATGTCCGCTCCCTC 4562
QY 3540 GTCTGTTCCCCAGGATCTCGAAGTGACTCCGGCTGAGCAGTGGGGCGGCTGGGGAGG 3599
Db 4563 GTCTGTTCCCCAGGATCTCGAAGTGACTCCGGCTGAGCAGTGGGGCGGCTGGGGAGG 4622
QY 3600 GGTGACGATTCTCCTCAGGCTTTGGCCCTGCAAGCAAAACCCACATATCTGTATGT 3659
Db 4623 GGTGACGATTCTCCTCAGGCTTTGGCCCTGCAAGCAAAACCCACATATCTGTATGT 4682
QY 3660 AATAAATGTCTTAACGTCG 3678
Db 4683 AATAAATGTCTTAACGTCG 4701
```

```
RESULT 5
AAI61066
ID AAI61066 standard; cDNA; 4956 BP.
XX
AC AAI61066;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 5055.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
```

```
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM41910.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
as central nervous system injuries.
XX
PS Claim 1; SEQ ID NO 5055; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 4956 BP; 1014 A; 1614 C; 1543 G; 785 T; 0 U; 0 Other;
Query Match 91.3%; Score 3374; DB 4; Length 4956;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 3478; Conservative 0; Mismatches 35; Indels 13; Gaps 7;
QY 180 GCCTCCCTGTCCTCGAGTGACTCCCAACGGGGCCAGGAGGACACGGGATGAGAA 239
Db 1194 GCCGCCAAAGTACAAAGCAGCTAGTCCCAACGGGGCCAGGAGGACACGGGATGAGAA 1253
QY 240 CGTGCCGGTGCCGGTGACTGCCGCCCTCTGGTGAGAGGAGGCCACCATGAAGCTGTG 299
Db 1254 CGTGCCGGTGCCGGTGACTGCCGCCCTCTGGTGAGAGGAGGCCACCATGAAGCTGTG 1313
QY 300 GTGTGCCCGGGGCTCAACCTGACGGGTGGAGGCCCAATGAGGACGACGCTGGGAATGG 359
Db 1314 GTGTGCCCGGGGCTCAACCTGACGGGTGGAGGCCCAATGAGGACGACGCTGGGAATGG 1373
QY 360 AGTCAAGCCAGCGCCAGGCCGCGATCCCCCTGACCTGCGACCGCGAGGAGACGGCGAGCC 419
Db 1374 AGTCAAGCCAGCGCCAGGCCGCGATCCCCCTGACCTGCGACCGCGAGGAGACGGCGAGCC 1433
QY 420 CAAGAGCGCCACACGCTCTCCGAGAAGAAGGCAAGGAGCTCCCTGAAATGGACGC 479
Db 1434 CAAGAGCGCCACACGCTCTCCGAGAAGAAGGCAAGGAGCTCCCTGAAATGGACGC 1493
QY 480 CACCTCCAGCGGGGTGGATCCTGACCAGCACCTGACCACCGAGAGGTGGTGATCAT 539
Db 1494 CACCTCCAGCGGGGTGGATCCTGACCAGCACCTGACCACCGAGAGGTGGTGATCAT 1553
QY 540 CGACGCCAAACAGCCGGGCACGGTGGTGGACCAAGTTACCGTCTGCAACGGCCACGTGCT 599
Db 1554 CGACGCCAAACAGCCGGGCACGGTGGTGGACCAAGTTACCGTCTGCAACGGCCACGTGCT 1613
QY 600 GTGCATCTCCAGCATCCCCCGGGCCAGCGACAGCGACTACCTCCCGGGGAGATGTTCTT 659
Db 1614 GTGCATCTCCAGCATCCCCCGGGCCAGCGACAGCGACTACCTCCCGGGGAGATGTTCTT 1673
```

QY 660 GGACAGCGACGTGAACCCAGAGGACCCCGGGCGCAGATGGCGTGTGGCGGTATCACCCCT 719  
Db |||||  
1674 GGACAGCGACGTGAACCCAGAGGACCCCGGGCGCAGATGGCGTGTGGCGGTATCACCCCT 1733  
QY 720 GGTGGGCTGTGCCACCCCGCTGCAACGTGCCGCGGAGCAACTGCTCTCTCCGAGGGGACAC 779  
Db |||||  
1734 GGTGGGCTGTGCCACCCCGCTGCAACGTGCCGCGGAGCAACTGCTCTCTCCGAGGGGACAC 1793  
QY 780 CCCAGTGTAGACAAGSGGCGAGGGGAGGTGGCCACCATCGCCAACGGGAAGGTCAACCC 839  
Db |||||  
1794 CCCAGTGTAGACAAGSGGCGAGGGGAGGTGGCCACCATCGCCAACGGGAAGGTCAACCC 1853  
QY 840 GTCCAGTCCACAGAGGAGGCCACAGAGGCCACCGAGGTGCCAGACCTGGGCCACGCGA 899  
Db |||||  
1854 GTCCAGTCCACAGAGGAGGCCACAGAGGCCACCGAGGTGCCAGACCTGGGCCACGCGA 1913  
QY 900 GCCAGAGACAGCCACATTGCGGCCCGGCCCTCTACAGAGCACTGTTCACTGACCCAGC 959  
Db |||||  
1914 GCCAGAGACAGCCACATTGCGGCCCGGCCCTCTACAGAGCACTGTTCACTGACCCAGC 1973  
QY 960 CCGACCCCGTCTCTTGGCCCCCAGCCTGGCAGCGAGAACGGGCCAGAGCCTGACAGCAG 1019  
Db |||||  
1974 CCGACCCCGTCTCTTGGCCCCCAGCCTGGCAGCGAGAACGGGCCAGAGCCTGACAGCAG 2033  
QY 1020 CAGCACACGGCCAGAGCCAGCCAGCGGGGACCCACGGGAGCAGGCAGCTGCTGC 1079  
Db |||||  
2034 CAGCACACGGCCAGAGCCAGAGCCAGCGGGGACCCACGGGAGCAGGCAGCTGCTGC 2093  
QY 1080 ACCCACCATTGTGGTGGGAGCCAGAACGGCTGGCTCTATGTGCACTGGCTGTGGCCAA 1139  
Db |||||  
2094 ACCCACCATTGTGGTGGGAGCCAGAACGGCTGGCTCTATGTGCACTGGCTGTGGCCAA 2153  
QY 1140 CTGGAAGAAGTGCCTGCACTCCATCAAGCTGAAGGATTCTGTGCTGAGCCTGGTGCATGT 1199  
Db |||||  
2154 CTGGAAGAAGTGCCTGCACTCCATCAAGCTGAAGGATTCTGTGCTGAGCCTGGTGCATGT 2213  
QY 1200 CAAAGGCCGTGTGCTGGTGGCTGTGGCGACGGGACCCCTGGCCATCTTCCACCGTGGTGA 1259  
Db |||||  
2214 CAAAGGCCGTGTGCTGGTGGCTGTGGCGACGGGACCCCTGGCCATCTTCCACCGTGGTGA 2273  
QY 1260 AGATGGCCAGTGGGATCTGAGCAACTATACCTAATGACCTGGGCCACCCGACCACTC 1319  
Db |||||  
2274 AGATGGCCAGTGGGATCTGAGCAACTATACCTAATGACCTGGGCCACCCGACCACTC 2333  
QY 1320 CATCCGCTGCATGGCTGTGTGTACGACCGCGTGTGGTGGCTACAGAACAAAGTGCA 1379  
Db |||||  
2334 CATCCGCTGCATGGCTGTGTGTACGACCGCGTGTGGTGGCTACAGAACAAAGTGCA 2393  
QY 1380 CGTCATCCAGCCCAAGACCATGCAGATAGAGAAGTCAATTGACGCCACCCGCGCGGGA 1439  
Db |||||  
2394 CGTCATCCAGCCCAAGACCATGCAGATAGAGAAGTCAATTGACGCCACCCGCGCGGGA 2453  
QY 1440 GAGCCAGTGGCGGAGTGGCGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCTGGA 1499  
Db |||||  
2454 GAGCCAGTGGCGGAGTGGCGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCTGGA 2513  
QY 1500 CTCACCCCTGAGGCTCTACCATGCACACGACCCAGCATCTACAGGACGTGGACATTGA 1559  
Db |||||  
2514 CTCACCCCTGAGGCTCTACCATGCACACGACCCAGCATCTACAGGACGTGGACATTGA 2573  
QY 1560 GCCCTACGTACGCAAGATGCTAGGCACTGGCAAGCTGGGTTCTCTCTTCGTACGCATCAG 1619  
Db |||||  
2574 GCCCTACGTACGCAAGATGCTAGGCACTGGCAAGCTGGGTTCTCTCTTCGTACGCATCAG 2633  
QY 1620 GGCCCTGCTTGTGCGGGCAGCCGGCTCTGGGTGGGACCCGGCAACGAGAGTGGTCACTC 1679  
Db |||||  
2634 GGCCCTGCTTGTGCGGGCAGCCGGCTCTGGGTGGGACCCGGCAACGAGAGTGGTCACTC 2693  
QY 1680 CATCCCCCTGACAGAGACTGTGGTCTCTGCACCCGAGGCA--GCTCTGGGGCTCCGAGCC 1737  
Db |||||  
2694 CATCCCCCTGACAGAGACTGTGGTCTCTGCACCGAGGCCAAGCTCTCTGGGGCTCCGAGCC 2753

QY 1738 AATAAGACATCCCCACCTCTTGGGG-AGGGCGCCCGCTCCCGGGGC-ATCATCCACGTGT 1795  
Db |||||  
2754 AATAAGACATCCCCACCTCTTGGGAAGGGCGCCCTCCCGGGGCAATCATCCACGTGT 2813  
QY 1796 AT-GGCGATGACAGAGTACAGGGCGGCCAGAGCTTTCATCCCCCTACTGCTCCATGGCC 1854  
Db |||||  
2814 ATGGCGATGACAGAGTACAGGGCGGCCAGAGCTTTCATCCCCCTACTGCTCCATGGCC 2873  
QY 1855 CAGGCCAGCTATGCTTCCATGGGCACCGCATGCCGTGAAGTTCTTGTCTCGGTGCCA 1914  
Db |||||  
2874 CAGGCCAGCTATGCTTCCATGGGCACCGCATGCCGTGAAGTTCTTGTCTCGGTGCCA 2933  
QY 1915 GGAACAGTGTGGCCACCCCTGAATGGCAGTGTGCTGACAGCCACGAGGCCCTGGG 1974  
Db |||||  
2934 GGAACAGTGTGGCCACCCCTGAATGGCAGTGTGCTGACAGCCACGAGGCCCTGGG 2993  
QY 1975 CCAGCTGCCCCCTGCTCGGAGGTGAGGGCCAGAAAGTGCAGAGCGTGGTGTGCTGAGC 2034  
Db |||||  
2994 CCAGCTGCCCCCTGCTCGGAGGTGAGGGCCAGAAAGTGCAGAGCGTGGTGTGCTGAGC 3053  
QY 2035 GGCGGGAGGGCTACATCGACTTTCGCATTGGAGACGGAGAGGACGAGACGGAGGAG 2094  
Db |||||  
3054 GGCGGGAGGGCTACATCGACTTTCGCATTGGAGACGGAGAGGACGAGACGGAGGAG 3113  
QY 2095 GGCGCAGGGGACATGAGCCAGGTGAAGCCCCGTGCTGTCCAAGGACAGCGCAGTCAATC 2154  
Db |||||  
3114 GGCGCAGGGGACATGAGCCAGGTGAAGCCCCGTGCTGTCCAAGGACAGCGCAGTCAATC 3173  
QY 2155 ATCGTGTGGCAGGTGTCTTACACCCCCGAGTGAAGCTGCTGCCCTGCCCTGGCCGACCTG 2214  
Db |||||  
3174 ATCGTGTGGCAGGTGTCTTACACCCCCGAGTGAAGCTGCTGCCCTGCCCTGGCCGACCTG 3233  
QY 2215 TACATAGGACCCCCGACCACTTGACCCCCCGCCCGGGGTAGCCAGCGCGCC 2274  
Db |||||  
3234 TACATAGGACCCCCGACCACTTGACCCCCCGCCCGGGGTAGCCAGCGCGCC 3293  
QY 2275 GCGGCCCTCTTTAACTCTCAACCTGACCTTTCACTGAGTCTGGCCCCCTCCAGCGG 2334  
Db |||||  
3294 GCGGCCCTCTTTAACTCTCAACCTGACCTTTCACTGAGTCTGGCCCCCTCCAGCGG 3353  
QY 2335 GCAGGAGTGGGGATGCGGATCAGCTGGGAGGAGGAGGGGTGCTTCCAGCCCCAGAGT 2394  
Db |||||  
3354 GCAGGAGTGGGGATGCGGATCAACTGGGAGGAGGAGGGAGGGGTGCTTCCACCCGA 3413  
QY 2395 GGGGAAGATGCTCTCGGACAGTTTCCCGGGCAGCTCTGGCCAGCTTCCAGCCCCAGAGT 2454  
Db |||||  
3414 GGGGAAGATGCTCTCGGACAGTTTCCCGGGCAGCTCTGGCCAGCTTCCAGCCCCAGAGT 3473  
QY 2455 CCTCAAGTCCAGGGACCTTGGGCCACGCGCAGGCAATCCGAGGTGGTCTTGCTCTA 2514  
Db |||||  
3474 CCTCAAGTCCAGGGACCTTGGGCCACGCGCAGGCAATCCGAGGTGGTCTTGCTCTA 3533  
QY 2515 CCCTGGGCTCTACTTCCCCAGCACCCCTGGAGGAGGAGGGGTCCCGCCGCGCGAGGC 2574  
Db |||||  
3534 CCCTGGGCTCTCTACTTCCCCAGCACCCCTGGAGGAGGAGGGGTCCCGCCGCGCGAGGC 3593  
QY 2575 TGCCTGCCCCGCGCCACCTCTGCACTGCTCATGGGGCCACCCCTGCTCTTGCGCCCT 2634  
Db |||||  
3594 TGCCTGCCCCGCGCCACCTCTGCACTGCTCATGGGGCCACCCCTGCTCTTGCGCCCT 3653  
QY 2635 CACTCTGCCTAGGGAGTGGGCCAGGCACTAGCCTTTGCCAGGAGGTGGCCCTCAGG 2694  
Db |||||  
3654 CACTCTGCCTAGGGAGTGGGCCAGGCACTAGCCTTTGCCAGGAGGTGGCCCTCAGG 3713  
QY 2695 CTGCCAGGTGCTGCACCCCAGCGGCTTCTCTGGGGCTTCCCGCTCGTCAAGCCTCT 2754  
Db |||||  
3714 CTGCCAGGTGCTGCACCCCAGCGGCTTCTCTGGGGCTTCCCGCTCGTCAAGCCTCT 3773  
QY 2755 ATCTGTCTGTCCCCACCCAGTGTCCCCCTGCCAGGAGTGGCATAAAGCAGGAGG 2814  
Db |||||  
3774 ATCTGTCTGTCCCCACCCAGTGTCCCCCTGCCAGGAGTGGCATAAAGCAGGAGG 3833  
QY 2815 CCGGCTCCCTGGGGCAGCTGTCTTGAGAACAGAGACTGTACCCCATCTCTGCCCATGCAG 2874

```
Db 3834 CCGGCTCCCTGGGCGAGCTGTGAGAACAGAGACTGTACCCCATCTGCCCATGCAG 3893
Qy 2875 GCAGGCTCTTCCAGCCCGTCTGTGACCCGCTGTCCCGCCAGGCTCTGCTGGGCGAGA 2934
Db 3894 GCAGGCTCTTCCAGCCCGTCTGTGACCCGCTGTCCCGCCAGGCTCTGCTGGGCGAGA 3953
Qy 2935 CTCACCTTGAGGAGTGGGCCCTGGAGTCTCTCCCTCCAGAACGCCCCCAGGCTGGAT 2994
Db 3954 CTCACCTTGAGGAGTGGGCCCTGGAGTCTCTCCCTCCAGAACGCCCCCAGGCTGGAT 4013
Qy 2995 TTCTCAGGCTGCCAGGGCAGGCCCTCAGGAAGAGGGGAGGCCCTGCGCTCTCC 3054
Db 4014 TTCTCAGGCTGCCAGGGCAGGCCCTCAGGAAGAGGGGAGGCCCTGCGCTCTCC 4073
Qy 3055 GGGATCAGTCTAGGACACAGGCTCAGCCTCAGGTTGATGGGGGATGATGTCTCCCGG 3114
Db 4074 GGGATCAGTCTAGGACACAGGCTCAGCCTCAGGTTGATGGGGGATGATGTCTCCCGG 4133
Qy 3115 GCCTGCTCTGACCGGGCTCCAGGAGCCAGCTCCAGACAGCTACTAAGTGCCTA 3174
Db 4134 GCCTGCTCTGACCGGGCTCCAGGAGCCAGCTCCAGACAGCTACTAAGTGCCTA 4193
Qy 3175 GGGTTGCCGCTGTGGCCTGTCTCCAGGGAGGACACAGAGAGGCCACCAAGCAGAGGCCCG 3234
Db 4194 GGGTTGCCGCTGTGGCCTGTCTCCAGGGAGGACACAGAGAGGCCACCAAGCAGAGGCCCG 4253
Qy 3235 TGGGGCTGAGGATGGAGCCGCCCGCAGCCGACTCAAGCCCGCAGAGGGCAGACGCCACC 3294
Db 4254 TGGGGCTGAGGATGGAGCCGCCCGCAGCCGACTCAAGCCCGCAGAGGGCAGACGCCACC 4313
Qy 3295 CTGACTGCTCTCCCTGCCAGCTGGGCTCTCTGGCC---TATCTCTACCTTCCAGGC 3350
Db 4314 CTGACTGCTCTCCCTGCCAGCTGGGCTCTCTGGCCCTAATTCTTAACCTTCCAGGC 4373
Qy 3351 CCACTGC-ACTCCTGTCTGGAGGCCCTTATGAGGGCAGCCC---AGCCCCCGACCCAC 3406
Db 4374 CCACTGCAACTCCTGTCTGGAGGCCCTTATGAGGGCAGCCCAGCCCCCGGACCCAC 4433
Qy 3407 CCCCAACAGAGAGACAGATCTTGGGGAGCTGCCCAAGCCCGCTGGCCACCCGAG 3466
Db 4434 CCCCAACAGAGAGACAGATCTTGGGGAGCTGCCCAAGCCCGCTGGCCACCCGAG 4493
Qy 3467 GGCTGAGCCGCTGGCTGCCGCTTCTCCCAACCCCTGCCACCTCCACTGTGATGTA 3526
Db 4494 GGCTGAGCCGCTGGCTGCCGCTTCTCCCAACCCCTGCCACCTCCACTGTGATGTA 4553
Qy 3527 TGTCCTCTCCCTCGTCTGTCTCCCGCAGGATCTCGAAGTGAATCTCCGGCTGAGCAGTGGG 3586
Db 4554 TGTCCTCTCCCTCGTCTGTCTCCCGCAGGATCTCGAAGTGAATCTCCGGCTGAGCAGTGGG 4613
Qy 3587 CGGCTGGGGAGGGGTGACGATCTCTCAGGCTTTGGCCCTGCAAGCAAAACCCACATAT 3646
Db 4614 CGGCTGGGGAGGGGTGACGATCTCTCAGGCTTTGGCCCTGCAAGCAAAACCCACATAT 4673
Qy 3647 CTGCTCTGTATGTAATAATGTCTTAACGTCTGTAATAAAAAAAAAA 3692
Db 4674 CTGCTCTGTATGTAATAATGTCTTAACGTCTGTAATAAAAAAAAAA 4719
```

RESULT 6  
AAS74091  
ID AAS74091 standard; cDNA; 5007 BP.

XX AAS74091;

AC AAS74091;

XX 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #9895.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.  
PN WO200175067-A2.  
XX 11-OCT-2001.  
PD 30-MAR-2001; 2001WO-US008631.  
XX 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX (HYSE-) HYSEQ INC.  
PA Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
DR P-PSDB; ABG09904.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX Claim 1; SEQ ID NO 9895; 103pp; English.  
PS The invention relates to isolated polynucleotide (I) and polypeptide (II)  
XX sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical disorders  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activities in  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 5007 BP; 1025 A; 1620 C; 1543 G; 805 T; 0 U; 14 Other;

Query Match 86.1%; Score 3182.2; DB 5; Length 5007;  
Best Local Similarity 97.0%; Pred. No. 0;  
Matches 3416; Conservative 0; Mismatches 83; Indels 24; Gaps 16;

Qy 180 GCCTCCCTGCTCCCTGCAGCTGAGTCCCAACGGGGCCAGGAGGACACGCGGATGAAGAA 239  
Db 1206 GCCCGCAAGTACAAGCAGCTGAGTCCCAACGGGGCCAGGAGGACACGCGGATGAAGAA 1265  
Qy 240 CGTGCCGCTGCCGCTGCTACTGCCGCC-TCTGGTGGAAGAGGACCCCAATGAAGCTGT 298  
Db 1266 CGTGCCGCTGCCGCTGCTACTGCCGCCCTTCTGTGGGAAAGACCCCACTGACCTGT 1325  
Qy 299 GGTGTGCCGCGGGCGTCAACCTGAGCGGTGGAGGCCCAATGAGGACGCTGGGAATG 358  
Db 1326 GCGTGCCCGCGCGCCACCTGAGCGGTGGAGGCCCAATGAGGACGCTGGGAATG 1385  
Qy 359 GAGTCAAGCCAGCGCCAGGCCCGC-GATCCCTGACCTGCGACCCGGAAGAGACGCGGAG 417  
Db 1386 GAGTCAAGCCAGCGCCAGGCCCGCCTGACCTGCGACCCGGAAGAGACGCGGAG 1445  
Qy 418 CCCAAGAGCGCCACACGCTCTCCCGAAGAAAGGCGCAAGGAGCTCCCTGAAATGGAC 477  
Db 1446 CCCAAGAGCGCCACACGCTCTCCCGAAGAAAGGCGCAAGGAGCTCCCTGAAATGGAC 1505

QY 478 GCCACCTCCAGCC-GGGTGTGGATCCTGACAGCACCCCTGACCACCAGCAAGGTGG-TGA 535  
Db |||||  
1506 GCCACCTCCAGCCGGGGTGTGGATCCTGACCAGCACCCCTGACCACCAGCAAGGTGGATGA 1565  
QY 536 TCATCGACGCCAAACAGCGGGGCACGGTGGTGGACCAAGTTCACCGTCTGCAACCGGCACG 595  
Db |||||  
1566 TCATCGACGCCAAACAGCGGGGCACGGTGGTGGACCAAGTTCACCGTCTGCAACCGGCACG 1625  
QY 596 TGCTGTGCATCTCCAGCATPCCCGCGGCAGCGACAGCGACTACCCCTCCCGGGGAGATGT 655  
Db |||||  
1626 TGCTGTGCATCTCCAGCATPCCCGCGGCAGCGACAGCGACTACCCCTCCCGGGGAGATGT 1685  
QY 656 TCCTGGACAGCGACGTGAACCCAGAGGACCCGGGCGCAGATGGCGTGTGCGCCGGTATCA 715  
Db |||||  
1686 TCCTGGACAGCGACGTGAACCCAGAGGACCCGGGCGCAGATGGCGTGTGCGCCGGTATCA 1745  
QY 716 CCCTGGTGGGTGTGCCACCCGCTGCAACGTTGCCGCGGAGCAACTGCTCCTCCCGAGGGG 775  
Db |||||  
1746 CCCTGGTGGGTGTGCCACCCGCTGCAACGTTGCCGCGGAGCAACTGCTCCTCCCGAGGGG 1805  
QY 776 ACACCCCAAGTGTGACAAAGGGGCAAGGGGAGGTGGCCACCATCGCCAAACGGGAAGGTCA 835  
Db |||||  
1806 ACACCCCAAGTGTGACAAAGGGGCAAGGGGAGGTGGCCACCATCGCCAAACGGGAAGGTCA 1865  
QY 836 ACCCGTCCAGTCCACAGAGGAGGCCACAGAGGCCACAGAGGTGGCGAGGAGGTGAC 895  
Db |||||  
1866 ACCCGTCCAGTCCACAGAGGAGGCCACAGAGGCCACAGAGGTGGCGAGGAGGTGAC 1925  
QY 896 GCGAGCCAGAGACAGCCACATTTGGCGCCCGGGCCCTCTCACAGAGCACGTCTTCACTGACC 955  
Db |||||  
1926 GCGAGCCAGAGACAGCCACATTTGGCGCCCGGGCCCTCTCACAGAGCACGTCTTCACTGACC 1985  
QY 956 CAGCCCCGACCCCGTCTCTGGCCCCCAGCCCTGGCAGCGAGGAGGTGCGACCCCTGAC 1015  
Db |||||  
1986 CAGCCCCGACCCCGTCTCTGGCCCCCAGCCCTGGCAGCGAGGAGGTGCGACCCCTGAC 2045  
QY 1016 GCAGCAGCACACGGCCAGAGCCAGAGCCAGCGGGGACCCCAACGGGAGCAGCAGTG 1075  
Db |||||  
2046 GCAGCAGCACACGGCCAGAGCCAGAGCCAGCGGGGACCCCAACGGGAGCAGCAGTG 2105  
QY 1076 CTGCACCCACCATGTGGCTGGAGCCCAAGACGGCTGGCTCTATGTGCACTCGGCTGTGG 1135  
Db |||||  
2106 CTGCACCCACCATGTGGCTGGAGCCCAAGACGGCTGGCTCTATGTGCACTCGGCTGTGG 2165  
QY 1136 CCAACTGGAAGAAGTGCCTGCACCTCCATCAAGCTGAAGGATTTCTGTGAGCCCTGGTGC 1195  
Db |||||  
2166 CCAACTGGAAGAAGTGCCTGCACCTCCATCAAGCTGAAGGATTTCTGTGAGCCCTGGTGC 2225  
QY 1196 ATGTCAAAGGCCGTGTGTGGTGGCTCTGCGGACGGGACCCCTGGCCATCTTCCACCGTG 1255  
Db |||||  
2226 ATGTCAAAGGCCGTGTGTGGTGGCTCTGCGGACGGGACCCCTGGCCATCTTCCACCGTG 2285  
QY 1256 GTGAAGATGGCCAGTGGGATCTGAGCAACTATCACCTAATGGAACCTGGGCCACCCGAC 1315  
Db |||||  
2286 GTGAAGATGGCCAGTGGGATCTGAGCAACTATCACCTAATGGAACCTGGGCCACCCGAC 2345  
QY 1316 ACTCCATCCGCTGCAATGGCTGTTGTGTACGACCCGCTGTGGTGTGGCTACAAGAACAGG 1375  
Db |||||  
2346 ACTCCATCCGCTGCAATGGCTGTTGTGTACGACCCGCTGTGGTGTGGCTACAAGAACAGG 2405  
QY 1376 TGACGCTCATCCAGCCCAAGACCATGAGATAGAGAGTCAATTTAGCCCCACCCGCGGC 1435  
Db |||||  
2406 TGACGCTCATCCAGCCCAAGACCATGAGATAGAGAGTCAATTTAGCCCCACCCGCGGC 2465  
QY 1436 GGGAGAGCCAGGTGCGCAGCTGGCGTGGATCGGCGATGGCGATGGGTGTCCATCCGCCC 1495  
Db |||||  
2466 GGGAGAGCCAGGTGCGCAGCTGGCGTGGATCGGCGATGGCGATGGGTGTCCATCCGCCC 2525  
QY 1496 TGGACTCCACCCCTGAGGCTCTACCATGACACACGACCCAGCATCTACAGGACGTGGACA 1555  
Db |||||  
2526 TGGACTCCACCCCTGAGGCTCTACCATGACACACGACCCAGCATCTACAGGACGTGGACA 2585  
QY 1556 TTGAGCCCTACGTACGCAAGATGCTAGGCACTGGCAAGCTGGGTTTCTCTTCGTACGCA 1615

Db |||||  
2586 TTGAGCCCTACGTACGCAAGATGCTAGGCACTGGCAAGCTGGGTTTCTCTCTCGTACGCA 2645  
QY 1616 TCACGGCCCTGCTTGTGCGGGCAGCCGGCTCTGGGTGGGCACCGGCAACGGAGTGGTCA 1675  
Db |||||  
2646 TCACGGCCCTGCTTGTGCGGGCAGCCGGCTCTGGGTGGGCACCGGCAACGGAGTGGTCA 2705  
QY 1676 TCTCCATPCCCTCTGAC-AGAGACTGTGGTCTGTGACCGAGGGCCA--GCTCCTGGGGCTCC 1732  
Db |||||  
2706 TCTCCATPCCCTCTGACAAAGAGACTGTGGTCTGTGACCGAGGGCCAAGCTCTGGGGCTCC 2765  
QY 1733 GAGCCAAATAAGACATCCCCCACTCTGGGG-AGGGGCCCGTCCCGGGGC-ATCATCCA 1790  
Db |||||  
2766 GAGCCAAATAAGACATCCCCCACTCTGGGGAAAGGGCCCGTCCCGGGGGCAATCATCCA 2825  
QY 1791 CGTGTAT-GGGATGACAGCAGTGAAGGGCGGCAGCAGCTTCAATCCCTACTGTCTCCA 1849  
Db |||||  
2826 CGTGTATGGGCGATGACAGCAGTGAAGGGCGGCAGAAAGCTTCAATCCCTACTGTCTCCA 2885  
QY 1850 TGGCCAGGCCCAAGCTATGCTTCCATGGGCACCGGATGCCGTGAAGTCTTGTGTCTCGG 1909  
Db |||||  
2886 TGGCCAGGCCCAAGCTATGCTTCCATGGGCACCGGATGCCGTGAAGTCTTGTGTCTCGG 2945  
QY 1910 TGCCAGGGAACTGTGTGGCCACCTGAATGGCAGTGTGTGGACAGCCCGAGGGGCC 1969  
Db |||||  
2946 TGCCAGGGAACTGTGTGGCCACCTGAATGGGAGTGTGTGGACAGCCCGAGGGGCC 3005  
QY 1970 CTGGGCCAGCTGCCCTCTCGGAGGTGAGGGCCAGAAAGCTGCGGAACCTGTGTGGTGC 2029  
Db |||||  
3006 CTGGGCCAGCTGCCCTCTCGGAGGTGAGGGCCAGAAAGCTGCGGAACCTGTGTGGTGC 3065  
QY 2030 TGAGCGCGGGGAGGGCTACATCGACTTCCGCAATGGAGACGAGAGGACGACGAGACGG 2089  
Db |||||  
3066 TGAGCGCGGGGAGGGCTACATCGACTTCCGCAATGGAGACGAGAGGACGACGAGACGG 3125  
QY 2090 AGGAGGGCGCAGGGGACATGAGCCAGGTGAAGCCCGTGTGTCTCCAAGGCAGAGCGCAGTC 2149  
Db |||||  
3126 AGGAGGGCGCAGGGGACATGAGCCAGGTGAAGCCCGTGTGTCTCCAAGGCAGAGCGCAGTC 3185  
QY 2150 ACATCATCGTGTGGCAGGTCTCTACACCCCGAGTGAAGTGTGTGCTGCCCTGGCCCCG 2209  
Db |||||  
3186 ACATCATCGTGTGGCAGGTCTCTACACCCCGAGTGAAGTGTGTGCTGCCCTGGCCCCG 3245  
QY 2210 ACCTGTACATAGGACCCCGACCCCTGACCCCGCCCGCCCGGGGTAGCCAGCCAG 2269  
Db |||||  
3246 ACCTGTACATAGGACCCCGACCCCTGACCCCGCCCGCCCGGGGTAGCCAGCCAG 3305  
QY 2270 GCGCCGCGCCCTTCTTAACCTCTCAACCTGACCTTTCACCTGAGTGTGGCCCCCTCC 2329  
Db |||||  
3306 GCGCCGCGCCCTTCTTAACCTCTCAACCTGACCTTTCACCTGAGTGTGGCCCCCTCC 3365  
QY 2330 AGCGGCAGGGAGTGGGGGATGCGGATCAACTGGGAGGAGGGGAGGGGTGCTTCCA 2389  
Db |||||  
3366 AGCGGCAGGGAGTGGGGGATGCGGATCAACTGGGAGGAGGGGAGGGGTGCTTCCA 3425  
QY 2390 CCCAGGGGAAGATGCTCTCGGACAGTTCCTCCCGGACAGTTCCTGGCCAGCTTCCAGCCC 2449  
Db |||||  
3426 CCCAGGGGAAGATGCTCTCGGACAGTTCCTCCCGGACAGTTCCTGGCCAGCTTCCAGCCC 3485  
QY 2450 AGAGTCTCAAGTCCAGGGCACCTTGGGCCACGCGCAGGCAGAAATCCGAGGTGGTCTGG 2509  
Db |||||  
3486 AGAGTCTCTAAGTCCAGGGCACCTTGGGCCACGCGCAGGCAGAAATCCGAGGTGGTCTGG 3545  
QY 2510 CTCTACCTTGGGCTCTACTCCCCAGCACCCCTGGAGGAGGAGGGGCTCCCCGCGGCC 2569  
Db |||||  
3546 CTCTACCTTGGGCTCTACTCCCCAGCACCCCTGGAGGAGGAGGGGCTCCCCGCGGCC 3605  
QY 2570 GAGGCTGCCTGCCTGGGCCCACTCTGCAATGTGCTCATGGGGCCACCTGCTCTCTGG 2629  
Db |||||  
3606 GAGGCTGCCTGCCTGGGCCCACTCTGCAATGTGCTCATGGGGCCACCTGCTCTCTGG 3665  
QY 2630 GCCTCACTCTGCCTAGGGAGTGGGCCAGGCACTAGCCTTTGCCACGGAGGTGGGCC 2689

|    |      |  |      |  |
|----|------|--|------|--|
| Db | 3666 | GCCTCACTCTGCCTAGGGAGCTGGGCCAGGCACCTAGCCTTTGCCAGGGAGGTGGCC    | 3725 | AAA62899 standard; DNA; 4173 BP.   |
| Qy | 2690 | TCAGGCTGCCAGGTGCTGCACCCAGCCGGCCTTCTCTGGGGCCTCCCGTCTGTC       | 2749 | AAA62899;  |
| Db | 3726 | TCAGGCTGCCAGGTGCTGCACCCAGCCGGCCTTCTCTGGGGCCTCCCGTCTGTC       | 3785 | 01-NOV-2000 (first entry)  |
| Qy | 2750 | CCTCTATCTCTGTCTCCACACCCAGCTGTCCCTGCCAGGGAGCTGGCATAAAGCA      | 2809 | Murine JNK3 binding protein DNA sequence.                                  |
| Db | 3786 | CCTCTATCTCTGTCTCCACACCCAGCTGTCCCTGCCAGGGAGCTGGCATAAAGCA      | 3845 | JNK3; Jun N-terminal protein kinase 3; JNK3 binding protein; JSAP;         |
| Qy | 2810 | CGAGGCCCGCTCCCTGGGGAGCTGCTTGAGAACAGAGACTGTACCCCATCTG-C       | 2868 | JNK/SAPK associated protein; nootropic; neuroprotective; neuroleptic;      |
| Db | 3846 | CGAGGCCCGCTCCCTGGGGAGCTGCTTGAGAACAGAGACTGTACCCCATCTGACC      | 3905 | antiparkinsonian; anticonvulsant; cerebroprotective; antiinflammatory;     |
| Qy | 2869 | ATGCAGGCAGGCTCTTGCCAGCCCGCTTCTGACCCCGTGTCCCGCCAGGCTCTGCTGGC  | 2928 | KW antiparkinsonian; anticonvulsant; cerebroprotective; antiinflammatory;  |
| Db | 3906 | ATGCAGGCAGGCTCTTGCCAGCCCGCTTCTGACCCCGTGTCCCGCCAGGCTTGTGCTGGC | 3965 | KW nervous system disease; Alzheimer's disease; Parkinson's disease;       |
| Qy | 2929 | AGAAGACTCACCTTGGAGGAGTGGGCCCTGGAGTCTCTCCCTCCAGAGCCCGCAGGG    | 2988 | KW Huntington's disease; multiple sclerosis; muscular dystrophy; apoplexy; |
| Db | 3966 | AGAAGACTCACCTTGGAGGAGTGGGCCCTGGAGTCTCTCCCTCCAGAGCCCGCAGGG    | 4025 | KW schizophrenia; dropsy; inflammatory disorders; mouse; ds.               |
| Qy | 2989 | TGGGATTTCTCAGGCTGCCAGGGCAGGCCCGCTCAGGAAGAGGGAGGCCCTGGC       | 3048 | XX Mus sp.   |
| Db | 4026 | TGGGATTTCTCAGGCTGCCAGGGCAGGCCCGCTCAGGAAGAGGGAGGCCCTGGC       | 4085 | XX Key   |
| Qy | 3049 | CTCTCCGGGATCAGTCTAGGACACAGGCTCAGCTCAGTTGATGGGGATGATGTCT      | 3108 | XX Location/Qualifiers   |
| Db | 4086 | CTCTCCGGGATCAGTCTAGGACACAGGCTCAGCTCAGTTGATGGGGATGATGTCT      | 4145 | XX 107..4024   |
| Qy | 3109 | CCGGGGCTGCTCTCTGACCGGGCTCCACGGAGCCAGTCTCCAGACACGCTACTAAG     | 3168 | XX /*tag= a  |
| Db | 4146 | CCGGGGCTGCTCTCTGACCGGGCTCCACGGAGCCAGTCTCCAGACACGCTACTAAG     | 4205 | XX /product= "JSAP"  |
| Qy | 3169 | TGCCTAGGGTTGCCGCTGTGGCTGTCTCCAGGGAGCAACAGAGAGGCCACCAAGCAGA   | 3228 | XX /note= "JNK3 binding protein"   |
| Db | 4206 | TGCCTAGGGTTGCCGCTGTGGCTGTCTCCAGGGAGCAACAGAGAGGCCACCAAGTGA    | 4265 | WO200031132-A1.  |
| Qy | 3229 | GGCCCGTGGGCTGAGATGGAGCCCGCCCGCTTATAGGGCAGCC--AGCCCCCGC       | 3288 | 02-JUN-2000.   |
| Db | 4266 | TGCCCGTGGGCTGAGATGGAGCCCGCCCGCTTATAGGGCAGCC--AGCCCCCGC       | 4325 | 19-NOV-1999; 99WO-JP006487.  |
| Qy | 3289 | GCCACCTGGACTGCTCTCCCTGCCAGTGGGCTCTCTGGCC---TATTCCTACCTT      | 3344 | 24-NOV-1998; 98JP-00332484.  |
| Db | 4326 | GCCACCTGGACTGCTCTCCCTGCCAGTGGGCTCTCTGGCC---TATTCCTACCTT      | 4385 | 02-SEP-1999; 99JP-00248442.  |
| Qy | 3345 | CCAGGCCCACTGC-ACTCTGTCTGGAGCCCTTATAGGGCAGCC--AGCCCCCGC       | 3400 | (KYOM ) KYOWA HAKKO KOGYO KK.  |
| Db | 4386 | CCAGGCCCACTGC-ACTCTGTCTGGAGCCCTTATAGGGCAGCC--AGCCCCCGC       | 4445 | Ichimura M, Hirose R, Yoshioka K;  |
| Qy | 3401 | ACCCACCCCAACAGAGAACACAGATCTTGGGAGCTGCCCCACAAG-CCCGCTGGC      | 3459 | WPI; 2000-400042/34.   |
| Db | 4446 | CACCACCCCAACAGAGAACACAGATCTTGGGAGCTGCCCCACAAGCCCGCTTGGC      | 4505 | P-PSDB; AAB12875.  |
| Qy | 3460 | CACCGA--GGGCTGCAGCCGCTGGCTGCC--GGCTTCTCCCGCCAGCCCTGCCACCTCC  | 3515 | Polypeptides binding to Jun N-terminal protein kinase for treatment and    |
| Db | 4506 | CACCGAAGGGCTTGCAATGCGTGGCTGCCCGGCTTCTCCCGACCCCTGCCACCTCC     | 4565 | diagnosis of nervous system and inflammatory disorders.                    |
| Qy | 3516 | ACTGTGATGATGTCCGCTCCCTGCTGTCTCCCGCAGGATCTCGAAGTACTCCGGCT     | 3575 | Disclosure; Page 110-123; 254pp; Japanese.                                 |
| Db | 4566 | ACTGTGATGATGTCCGCTCCCTGCTGTCTCCCGCAGGATCTCGAAGTACTCCGGCT     | 4625 | The invention relates to polypeptides which bind to JNK3 (Jun N-terminal   |
| Qy | 3576 | GAGCAGTGGGGCGGCTGGGGAGGGGTGACGATCTCTCAGGCTTTGGCCCTGCAAGCA    | 3635 | protein kinase 3). The present sequence represents a DNA sequence          |
| Db | 4626 | GAGCAGTGGGGCGGCTGGGGAGGGGTGACGATCTCTCAGGCTTTGGCCCTGCAAGCA    | 4685 | encoding a JNK3 binding protein (JSAP; JNK/SAPK associated protein).       |
| Qy | 3636 | AACCCACATATCTGTATGTAATAAATGCTTTAACGTG 3678                   |      | Including in the invention are polypeptides derived from the JNK3 binding  |
| Db | 4686 | AACCCCAATATTGCTCTGTATGTAATAAATGCTTTAACGTG 4728               |      | proteins, DNA sequences encoding the proteins, expression vectors          |
|    |      |  |      | containing the DNA, and host cells transformed by the vectors. The         |
|    |      |  |      | polypeptides exhibit nootropic, neuroprotective, antiparkinsonian,         |
|    |      |  |      | anticonvulsant, cerebroprotective, neuroleptic, and antiinflammatory       |
|    |      |  |      | activity. JNK3 binding proteins are used in the treatment, prevention and  |
|    |      |  |      | diagnosis of diseases associated with the JNK3 cascade, such as disease,   |
|    |      |  |      | of the nervous system including Alzheimer's disease, Parkinson's disease,  |
|    |      |  |      | Huntington's disease, multiple sclerosis, muscular dystrophy, apoplexy     |
|    |      |  |      | and schizophrenia, dropsy, and inflammatory disorders                      |
|    |      |  |      | XX Sequence 4173 BP; 1027 A; 1146 C; 1214 G; 786 T; 0 U; 0 Other;          |
|    |      |  |      | SQ   |
|    |      |  |      | Query Match 41.5%; Score 1531.6; DB 3; Length 4173;                        |
|    |      |  |      | Best Local Similarity 82.3%; Pred. No. 2.9e-289;                           |
|    |      |  |      | Matches 1785; Conservative 0; Mismatches 379; Indels 6; Gaps 2;            |
| Qy | 148  | GATGGGTAGAGCCAGGGTTCGTGCCCGCCCTCCCTGCTCCCTGCCAGTGA           | 207  |  |
| Db | 1985 | GACGGAGGCTGCAGGCCCTGTGGTGGAGCTGCCAAGTACAAGCAGTGA             | 2044 |  |
| Qy | 208  | AACGGGGCCAGGAGGACACGGGATGAAGAACGTGCCGGTGCCTGCTG              | 267  |  |
| Db | 2045 | AATGAGGCCAGGAAGACACCGGATGAAAATGTGCTGCTCCCTGTG                | 2104 |  |
| Qy | 268  | CTGGTGGAGAGGACCCACCATGAAGCTGTGGTGTGCCCGGGCGTCAAC             | 327  |  |



DT 01-NOV-2000 (first entry)  
XX Murine JNK3 binding protein DNA sequence #2.  
DE JNK3; Jun N-terminal protein kinase 3; JNK3 binding protein; JSAP;  
XX JNK/SAPK associated protein; nootropic; neuroprotective; antiepileptic;  
KW antiparkinsonian; anticonvulsant; cerebroprotective; antiinflammatory;  
KW nervous system disease; Alzheimer's disease; Parkinson's disease;  
KW Huntington's disease; multiple sclerosis; muscular dystrophy; apoplexy;  
KW schizophrenia; drowsy; inflammatory disorders; mouse; ds.  
XX Mus sp.  
OS  
XX  
XX Key Location/Qualifiers  
FH 107..4052  
FT /\*tag= a  
FT /product= "JSAP"  
FT /note= "JNK3 binding protein"  
XX  
XX WO200031132-A1.  
XX  
XX 02-JUN-2000.  
XX  
XX 19-NOV-1999; 99WO-JP006487.  
XX  
XX 24-NOV-1998; 98JP-00332484.  
XX  
XX 02-SEP-1999; 99JP-00248442.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
XX Ichimura M, Hirose R, Yoshioka K;  
PI WPI; 2000-400042/34.  
XX P-PSDB; AAB12876.  
XX  
XX Polypeptides binding to Jun N-terminal protein kinase for treatment and  
PT diagnosis of nervous system and inflammatory disorders.  
PT  
XX  
XX Claim 4; Page 123-136; 254pp; Japanese.  
XX  
XX The invention relates to polypeptides which bind to JNK3 (Jun N-terminal  
CC protein kinase 3). The present sequence represents a DNA sequence  
CC encoding a JNK3 binding protein (JSAP; JNK/SAPK associated protein).  
CC Included in the invention are polypeptides derived from the JNK3 binding  
CC proteins, DNA sequences encoding the proteins, expression vectors  
CC containing the DNA, and host cells transformed by the vectors. The  
CC polypeptides exhibit nootropic, neuroprotective, antiparkinsonian,  
CC anticonvulsant, cerebroprotective, neuroleptic, and antiinflammatory  
CC activity. JNK3 binding proteins are used in the treatment, prevention and  
CC diagnosis of diseases associated with the JNK3 cascade, such as diseases  
CC of the nervous system including Alzheimer's disease, Parkinson's disease,  
CC Huntington's disease, multiple sclerosis, muscular dystrophy, apoplexy  
CC and schizophrenia, drowsy, and inflammatory disorders  
XX  
XX Sequence 4200 BP; 1034 A; 1153 C; 1223 G; 790 T; 0 U; 0 Other;  
SQ  
Query Match 41.5%; Score 1531.6; DB 3; Length 4200;  
Best Local Similarity 82.3%; Pred. No. 2.9e-289;  
Matches 1785; Conservative 0; Mismatches 379; Indels 6; Gaps 2;  
QY 148 GATGGGTAGGAGCCAGGGTGTGTCGCCACGGCGCCCTCCCTGCTCCCTGAGCTGAGTCCC 207  
Db 2012 GACGGGAGGCTGCAGGCCCTGTGGGTGGAGCCTGCCTGCCAAGTACAAGCAGCTGAGCCCC 2071  
QY 208 AACGGGGCCAGGAGGACACGGGGATGAAGAACGTGCCGGTCCGGTGTACTGCCGCCCT 267  
Db 2072 AATGGAGGCCAGGAAGACACCGGGATGAATAATGTGCCTGTCCCTGTGTACTGTGCCCT 2131  
QY 268 CTGGTGGAGAGGACCCCATGAAGCTGTGGTGTGCCGGGGCGTCAACCTGAGCGGG 327  
Db 2132 CTGGTGGAGAGGACCCCTTCGACAAAGCTGTGGTGTGCTGTGTGTCAACCTGAGTGGG 2191  
QY 328 TGGAGGCCCAATGAGGACGACGCTGGGAATGAGTCAAGCCAGGCCGCGGATCCC 387

Db 2192 TGAAGCCACATGAAGAGGACTCTAGCAATGGACCCAGCCTGTACCAGTCGAGACCCT 2251  
QY 388 CTGACCTGCGACCGCGAAGAGAGACGGCGAGCCCAAGAGCGCCACACAGTCTCTCCGAGAAG 447  
Db 2252 CTGACCTGTGACCGGGAAGGAGAGCGGAACCCAGAGCACACACCCATCACCTGAGAAG 2311  
QY 448 AAGAAGGCCAAGGAGCTCCCTGAAATGAGACGCCACCTCCAGCCGGGTGTGGATCCTGACC 507  
Db 2312 AAGAAGGCCAAGGAAACCCCTGAGGAGAGTGTACCTCCAGTCGGGTATGGATCCTCACC 2371  
QY 508 AGCACCTTGACCAACAGCAAGGTGGTGTATCATCGACGCCAACCCAGCCGGGACGGTGGTG 567  
Db 2372 AGCACCTTGACAACAGCAAGGTGGTGTATCATTTGATGCCAACCCAGCCAGGCAATTTGTG 2431  
QY 568 GACCAGTTCAACCGTCTGCAACGCGCACGTCGTGTGTCATCTCCAGCATCCCGCGGCCAGC 627  
Db 2432 GATCAGTTCAAGTCTGCAATGCCACGTCCTGTGTATCTCCAGCATTCCTGCGGCCAGT 2491  
QY 628 GACAGCGACTACCCCTCCCGGGAGAGATGTTCTTGGACAGCGACGTGAACCCAGAGGACCCG 687  
Db 2492 GACAGTGACTATCCCTCGGGAGAGATGTTCTTAGACAGTGTGAACCCCTGAAGATTCA 2551  
QY 688 GCGCGAGATGGGTGTGTCGCGGTATCACCTGGTGGGTGTGTCACCCCGTCCAGCCGTCACACGTG 747  
Db 2552 GGTGCTGATGGTGTGTCGCTGGCATCACCTGGTGGGTGTGCTACCCGCTGCAATGTT 2611  
QY 748 CCGCGGAGCAACTGCTCTCCCGAGGGGACACCCCGTGTGCTAGACAAGGGGAGGGGGAG 807  
Db 2612 CCACGTAGCAACTGTTCTCACGAGGAGACACCCCGTGTGCTAGCAAGGGGAGGGGGAT 2671  
QY 808 GTGGCCACCATCGCCAAACGGGAAGGTCAACCCCTGCCAGTCCACAGAGGAGGCCACAGAG 867  
Db 2672 GTGGCGACCATGCTCAATGGGAAGGTCAACCCCTGCCAATCCACAGAAGAGGCCACAGAA 2731  
QY 868 GCCACGGAGGTGCCAGACCCCTGGGCCAGCGAGCAGACAGACAGCAGCATTGCGGCCCGGG 927  
Db 2732 GCCACAGAGGTGCCAGACCCCTGTCGCCAGCGAGTCAAGAGCAACGACAGTCCGCGCCGGG 2791  
QY 928 CCTCTCACAGAGCAGCTTCTACTGACCCAGCCCGACCCCGTCTCTGGCCCCCAGCCT 987  
Db 2792 CCTCTCACAGAGCAGTCTTTACTGACCCAGCACCCACCCCATCTCTCCAGCACCCAGCCT 2851  
QY 988 GGCACGGAGAACGGGCCAGAGCCTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1047  
Db 2852 GCCAGTGAGAAATGGGTGAGAATCCCAATGGCACCATTGTACAGCCTCAGGTGGAGCCAGT 2911  
QY 1048 GGGGACCCACGGGAGCAGGAGCAGTGTGCACCCACCATGTGGTGGGAGCCAGAAC 1107  
Db 2912 GGGGAACTCTCAACAAACACAGTAGCGTGCACCCACTATGTGGTAGGAGCCAGAAAT 2971  
QY 1108 GGCTGGCTCTATGTGCACTCGGCTGTGGCCAACTGGAAGAAGTGCCTGCACCTCCATCAAG 1167  
Db 2972 GGCTGGCTCTATGTGCATTGACGGGTAGCCAACTGGAAGAAGTGTCTGCACCTCCATCAAG 3031  
QY 1168 CTGAAGGATTCTGTGCTGAGCCTGGTGCATGTCAAAGGCCGTGTGCTGGTGGTCTTGGCG 1227  
Db 3032 CTAAAGACTCTGTGTGAGCCTGGTGCATGTCAAAGGCCGAGTGTGCTGGTAGTCTTTGCA 3091  
QY 1228 GACGGGACCTGGCCATCTTCCACCGTGGTGAAGATGGCAGTGGGATCTGAGCAACTAT 1287  
Db 3092 GATGGGACCTGGCTATCTTCCATCGTGGAGAGGATGGCCAGTGGGACCTGAGCAACTAC 3151  
QY 1288 CACCTAATGGACCTGGGCCACCCGACCCACTCCATCCGCTGCATGGCTGTGTGTGTACGAC 1347  
Db 3152 CACCTAATGGACCTGGGCCACCCACACCACTCCATCCGCTGCATGGCTGTGTGTGTGTGATGAC 3211  
QY 1348 CGCGTGTGGTGTGGCTACAAGAACAAGGTGCACGTTCATCCAGCCCAAGACCATGCAGATA 1407  
Db 3212 CGAGTTTGGTGTGGCTACAAGAACAAGGTGCATGTTATCCAGCCCAAGACAATGCAGATT 3271  
QY 1408 GAGAAGTCATTGACGCCCAACCCGCGCGGAGAGCCAGTGCAGCTGGCGTGGGATC 1467

Db 3272 GAGAAATCATTTGATGCCACCAAGCGGGGAAAGCCAGGTACGTACGTGGCCTGGATC 3331  
QY 1468 GGCAGTGGCGTATGGGTGTCATCCGCTGGAGACTCCACCTGAGGCTTACCATGCACAC 1527  
Db 3332 GGTGATGAGTGTGGGTCTCTATTTCGCTGGATTCTACCTTCGGCTTACCATGCTCAC 3391  
QY 1528 ACGCACAGCATCTACAGGACGTGGACATGAGCCCTACGTGAGCAAGATGCTAGGCACT 1587  
Db 3392 ACCACAGCACCTGCAGGATGTGGACATGAGCCCTATGTTAGCAAGATGCTAGGAACC 3451  
QY 1588 GGCAAGTGGGTTTCTCTCTTCGTACGCATCAGGCGCTGCTTGTGCGGGGAGCCGCTC 1647  
Db 3452 GGCAAGTGGGCTTCTCTCTTCGTGCGCATCACAGCCTTACTCATTTGAGGCAACCGTCTG 3511  
QY 1648 TGGTGGGCACCGGCAACGGAGTGTTCATCTCCATCCCTGACAGAGACTGTGGTCTCTG 1707  
Db 3512 TGGTGGGCACCTGGCAATGGGTTGTTCATCTCCATCCCTTGACTGAGACTGTGGTCTCTG 3571  
QY 1708 CACGAGGCCAGCTCCTGGGCTCCGAGCCCAATAGACATCCCCACCTCTTGGGAGGGC 1767  
Db 3572 CATGAGGCCAGCTCCTAGGCTCCGAGCCCAACAGACATCCCCAACATCTGGGAGGGG 3631  
QY 1768 GCGCGTCCCGGGGATCATCCACGTGTATGGCGATGACAGAGTGACAGGGCGGCCAGC 1827  
Db 3632 ACCGCGCCAGGGGCATCATCCATGTGTATGGGACGACAGCAGTGACAGGCCGCCAGT 3691  
QY 1828 AGCTTCATCCCTTACTGCTCCATGGCCAGGCCAGCTATGCTTCCATGGGCACCGCAT 1887  
Db 3692 AGTTTCATCCCTTACTGCTCCATGGCCAGGCCAGCTTGTCTCCATGGGCACCGTAT 3751  
QY 1888 GCCGTGAAGTTCTTTGTCTCGGTGCCAGGAACTGCTGGCCACCTTGAATGGCAGTGTG 1947  
Db 3752 GCTGTCAAATCTTTGTCTCTGTGCCAGGAATGTGTGGCCACTCTCAATGGCAGTGTG 3811  
QY 1948 CTGGACAGCCAGCCGAGGGCCCTGGCCAGCTGCCCCCTGCTCGGAGTCCGAGGGCCAG 2007  
Db 3812 CTAGACAGCCCATCAGAGGGCCCTGGCCCTGCTGCACCCGCTGCAGATGCTGAGGGCCAG 3871  
QY 2008 AAGCTGCGGAACGTGTGTGCTGAGCGGCGGGGAGGGCTACATCGACTTCGCGATTGGA 2067  
Db 3872 AAGTTGAAGATGCACTGCTGTGAGTGTGTGAGTGTGAGGTTACATTGACTTCGATCGGA 3931  
QY 2068 GACGGAGAGCAGCAGACGAGACGAGGAGGGCGCAGGGGACATGAGCCAGGTGAAGCCCGTG 2127  
Db 3932 GACGGAGAGGATGATGAACATGAGGAATGTGCGGGGACGTGAACAGACAAAGCCCTCG 3991  
QY 2128 CTGTCCAGCAGAGCGCAGTCACATCATCTGTGTGGCAGGTGTCTTACACCCCGAGTGA 2187  
Db 3992 TTGTCCAGGCTGAGCGCAGCCACATCATCTGTGTGGCAGGTGTCTTACACCCCTGAGTG- 4050  
QY 2188 AGCTGTGCTGCTGCTGCGCCGACCTGTACATAGGACCCCGACCATCTGACCCCGCCC 2247  
Db 4051 AGACCTGTCTTACCTGATGCCAATGTACATAGGACCCCTACCTGCTGCTGCTGCTGCTGCT 4110  
QY 2248 GGCCCGGGGTAGCCAGCGCGCGCCGCGCCCTCTTCTTAACCTCTCAACCTGCGAGCT 2307  
Db 4111 GTTCCCTGGGCAGCCA-----GGTTCTGCTCATCCCTTTTAACTCTCAACTTGCAGCT 4165  
QY 2308 TTCACCTGAG 2317  
Db 4166 TTTGCCTGAG 4175

RESULT 9  
AAA62902

ID AA62902 standard; DNA; 4266 BP.

XX  
AC AAA62902;

XX  
DT 01-NOV-2000 (first entry)

XX  
DE Murine JNK3 binding protein DNA sequence #4.

KW JNK3; Jun N-terminal protein kinase 3; JNK3 binding protein; JSAP;  
KW JNK/SAPK associated protein; nootropic; neuroprotective; neuropleptic;  
KW antiparkinsonian; anticonvulsant; cerebroprotective; antiinflammatory;  
KW nervous system disease; Alzheimer's disease; Parkinson's disease;  
KW Huntington's disease; multiple sclerosis; muscular dystrophy; apoplexy;  
KW schizophrenia; dropsy; inflammatory disorders; mouse; ds.  
XX Mus sp.

Key Location/Qualifiers  
CDS 107..4117  
/\*tag= a  
/product= "JSAP"  
/note= "JNK3 binding protein"

WO200031132-A1.

02-JUN-2000.

19-NOV-1999; 99WO-JP006487.

24-NOV-1998; 98JP-00332484.

02-SEP-1999; 99JP-00248442.

(KYOW ) KYOWA HAKKO KOGYO KK.

Ichimura M, Hirose R, Yoshioka K;

WPI; 2000-400042/34.

P-PSDB; AAB12878.

Polypeptides binding to Jun N-terminal protein kinase for treatment and  
diagnosis of nervous system and inflammatory disorders.

Claim 4; Page 150-164; 254pp; Japanese.

The invention relates to polypeptides which bind to JNK3 (Jun N-terminal  
protein kinase 3). The present sequence represents a DNA sequence  
encoding a JNK3 binding protein (JSAP; JNK/SAPK associated protein).  
Included in the invention are polypeptides derived from the JNK3 binding  
proteins, DNA sequences encoding the proteins, expression vectors  
containing the DNA, and host cells transformed by the vectors. The  
polypeptides exhibit nootropic, neuroprotective, antiparkinsonian,  
anticonvulsant, cerebroprotective, neuropleptic, and antiinflammatory  
activity. JNK3 binding proteins are used in the treatment, prevention and  
diagnosis of diseases associated with the JNK3 cascade, such as diseases  
of the nervous system including Alzheimer's disease, Parkinson's disease,  
Huntington's disease, multiple sclerosis, muscular dystrophy, apoplexy  
and schizophrenia, dropsy, and inflammatory disorders

Sequence 4266 BP; 1045 A; 1176 C; 1245 G; 800 T; 0 U; 0 Other;

Query Match 41.5%; Score 1531.6; DB 3; Length 4266;  
Best Local Similarity 82.3%; Pred. No. 2.9e-289;  
Matches 1785; Conservative 0; Mismatches 379; Indels 6; Gaps 2;

QY 148 GATGGTAGGAGCCAGGTTGTTGTCGCCACGGCGCTCCCTGCTCCCTGAGTCC 207  
Db 2078 GACGGAGGCTGACGCTGTGGTGGAGCCTGCTGCCAAGTACAGCAGCTGAGCCC 2137  
QY 208 AACGGGGCCAGGAGGACACGGGATGAAGAACGTGCGGTGCCGTGCTGCTGCTGCT 267  
Db 2138 AATGGAGGCCAGGAGACACCGGATGAAAAATGTGCTGCTGCTGCTGCTGCTGCT 2197  
QY 268 CTGTTGGAGAGGAGCCACCATGAAGCTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCT 327  
Db 2198 CTGTTGGAGAGGAGCCCTTCGACAAAGCTGTGTGTGCTGCTGCTGCTGCTGCTGCT 2257  
QY 328 TGGAGGCCCAATGAGGACGACGCTGGGAATGAGTCAAGCCAGCCAGCCGCGATCCC 387  
Db 2258 TGGAGGCCCAATGAGGAGGACTTAGCAATGGACCCAGGCTGTGCTGCTGCTGCTGCT 2317  
QY 388 CTGACCTGCGACCGGAGGAGAGAGCGGAGGCCAAGAGCGCCACACGCTCTCCGAGAAG 447

|    |      |  |      |
|----|------|--|------|
| Db | 2318 | CTGACCTGTGACCGGAAGAGAAAGCGGAACCCCAAGAGCACACACCCATCACCTGAGAAAG  | 2377 |
| QY | 448  | AAGAAGGCCAAGGAGCTCCTCGAAATGGACGCCACCTCCAGCCGGGTGTGGATCCTGACC   | 507  |
| Db | 2378 | AAGAAGGCAAGGAACCCCTGAGGCAGATGCTACCTCCAGTCGGGTATGGATCCTCACC     | 2437 |
| QY | 508  | AGCACCTTGACCAACAGCAAGGTGGTGATATCGACGCCCAACCAAGCCGGGCACGGTGGTG  | 567  |
| Db | 2438 | AGCACCTTGACCAACAGCAAGGTGGTGATATCGACGCCCAACCAAGCCGGGCACAAATTGTG | 2497 |
| QY | 568  | GACCAATTTCACCGTCTGCAACCGGCACGTGCTGTGCATCTCCAGCATCCCCGGGCCAGC   | 627  |
| Db | 2498 | GATCAGTTTCACAGTCTGCAATGCCACAGTCCCTGTGTATCTCCAGCATTCCTGCGGCCAGT | 2557 |
| QY | 628  | GACAGCGACTACCTTCCCGGGAGATGTTCTTGGAACAGCGACCTGAACCCAGAGGACCCG   | 687  |
| Db | 2558 | GACAGTGACTATCCCCCTGGGAGATGTTCTTAGACAGTGTGAACCCCTGAAGATTCA      | 2617 |
| QY | 688  | GGCGCAGATGGCGTCTGCGCCGGTATCACCTGTGTGGTGTGCGCACCCGCTGCAACGTG    | 747  |
| Db | 2618 | GGTGTGATGGTGTGCTGGCTGGCATCACCTGGTGGGTGTGCTACCCGCTGCAATGTT      | 2677 |
| QY | 748  | CGCGGAGCAACTGCTCCTCCGAGGGGACACCCAGTGTAGCAAGGGGCAGGGGGAG        | 807  |
| Db | 2678 | CCACGTAGCAACTGTTCTCTACGAGGAGACACCCAGTACTGGACAAGGGGCAGGGGGAT    | 2737 |
| QY | 808  | GTGGCCACCATCGCCAAACGGGAAGGTCAACCCGTCCAGTCCACAGAGGAGGCCACAGAG   | 867  |
| Db | 2738 | GTGGGACCACTGCGCAATGGGAAGGTCAACCCGTCCCAATCCACAGAAGGCCACAGAA     | 2797 |
| QY | 868  | GCCACGGAGGTGCCAGACCCTGGGCCAGCGAGCCAGAGACAGCCACATTTGGGCCCGGG    | 927  |
| Db | 2798 | GCCACAGAGGTGCCAGACCCTGTTCCACGCGAGTCAGAAAGCAACGACAGTCCGGGCCGGG  | 2857 |
| QY | 928  | CCTCTCACAGAGCAGTCTTCACTGACCCAGCCCGACCCCGTCTCTGCGCCCGCAGCCT     | 987  |
| Db | 2858 | CCTCTCACAGAGCAGTCTTCTTACTGACCCAGCACCCACCCCATCTCTCCAGCACCCAGCCT | 2917 |
| QY | 988  | GGCAGCGAGAACGGGCAGAGCCTTGACAGCAGCAGCACACGGCCAGAGCCAGGCCAGC     | 1047 |
| Db | 2918 | GCCAGTGAGAATGGGTGAGAATCCAATGGCACCAATTGTACAGCCTCAGGTGGAGCCCAGT  | 2977 |
| QY | 1048 | GGGGACCCACGGGAGCAGGCAGTGTGACCCACCATGTGGCTGGGAGCCCGAGAAC        | 1107 |
| Db | 2978 | GGGGAATCTCAACAACAACCAAGTAGCGCTGCACCCCACTATGTGGCTAGGAGCCCAAGAT  | 3037 |
| QY | 1108 | GGCTGGCTCTATGTGCACTCGGCTGTGGCCAACTGGAGAAGTGCCTGCATCCATCAAG     | 1167 |
| Db | 3038 | GGCTGGCTCTATGTGCATTGACGGTAGCCCACTGGAAGAAAGTGTCTGCACTCCATCAAG   | 3097 |
| QY | 1168 | CTGAAGGATTCTGTGTGAGCCTGGTGATGTCAAAAGCCGTGTGCTGGTGGCTTGGCG      | 1227 |
| Db | 3098 | CTAAAGACTCTGTGTGAGCCTGGTGATGTCAAAAGCCGAGTGTCTGGTAGCTCTTGCA     | 3157 |
| QY | 1228 | GACGGACCCCTGGCCATCTTCCACCGTGGTGAAGATGGCCAGTGGGATCTGAGCAACTAT   | 1287 |
| Db | 3158 | GATGGGACCCCTGGCTATCTTCCATCGTGGAGAGGATGGCCAGTGGGACCTGAGCAACTAC  | 3217 |
| QY | 1288 | CACCTAATGACCTGGGCCACCCCGCACCACTCCATCCGCTGCATGGCTGTGTGTACGAC    | 1347 |
| Db | 3218 | CACCTAATGACCTGGGCCACCCACACCACTCCATCCGCTGCATGGCTGTGTGAATGAC     | 3277 |
| QY | 1348 | CGCGTGTGGTGTGGCTACAAGAAACAAGGTGCACGTATCCAGCCCCAAGACCATGCAGATA  | 1407 |
| Db | 3278 | CGAGTTTGGTGTGGCTACAAGAAACAAGGTGCATGTTATCCAGCCCCAAGACAATGCAGATT | 3337 |
| QY | 1408 | GAGAAATCAATTTGACGCCACCCCGCGGGAGAGCCAGTGGCGGACGTGGCGTGGATC      | 1467 |
| Db | 3338 | GAGAAATCAATTTGATGCCACCCCAAGCGGGGAAAGCCAGGTACGTGAGCTGGCCTGGATC  | 3397 |
| QY | 1468 | GGCGATGGCGTATGGGTGTCATCCGCTGGAATCCACCCCTGAGGCTCTACCATGCACAC    | 1527 |

|    |      |   |      |
|----|------|---|------|
| Db | 3398 | GGTGATGAGTGTGGGTCTCTATTTCGCTTGGATTCTACCCCTTCGGCTTACCATTGCTCAC     | 3457 |
| QY | 1528 | ACGCACCAAGCATCTACAGACGTGGACATTGAGCCCTACGTTCAGCAAGATGCTAGGCAT      | 1587 |
| Db | 3458 | ACCCACCAAGCATCTGCAGGATGTGGACATTGAGCCCTATGTTAGCAAGATGCTAGGAACC     | 3517 |
| QY | 1588 | GGCAAGCTGGGTTTCTCCTTCGTACGCATACGGCCCTGCTTGTGCGGGCAGCCGGCTC        | 1647 |
| Db | 3518 | GGCAAGCTGGGCTTCTCCTTCGTGCGCATCACAGCCTTACTCATTTGCAGGCAACCGTCTG     | 3577 |
| QY | 1648 | TGGTGGGCACCGCAACCGAGTGGTCATCTCCATCCCGCTGACAGAGACTTGGTCCCTG        | 1707 |
| Db | 3578 | TGGGTGGGCACTGGCAATGGGGTGTTCATCTCCATCCCGCTGACTGAGACTTGGTCCCTG      | 3637 |
| QY | 1708 | CACGAGGCCAGCTCCTGGGCTCCGAGCCCAATAAGACATCCCCACCTCTGGGGAGGGC        | 1767 |
| Db | 3638 | CATCGAGGCCAGCTCCTAGGCTCCGAGCCCAACAAGACATCCCCAACATCTGGGGAGGGG      | 3697 |
| QY | 1768 | GCCCGTCCCGGGGCATCATCCACGTGTATGGCGATGACAGAGTGACAGGGCGGCCAGC        | 1827 |
| Db | 3698 | ACCCGCCAGGGGCATCATCCATGTGTATGGGACGACAGAGTGACAGGGCCGACGT           | 3757 |
| QY | 1828 | AGCTTCATCCCGTACTGCTCCATGGCCAGGCCAGCTATGCTTCCATGGGCACCGCGAT        | 1887 |
| Db | 3758 | AGTTTCATCCCGTACTGCTCCATGGCACAGGCTCAGCTTTCCTTCCATGGGCACCGTAT       | 3817 |
| QY | 1888 | GCCGTGAAGTTCTTGTCTCGGTGCCAGGGAAGTGTCTGGCCACCTGAATGGCAGTGTG        | 1947 |
| Db | 3818 | GCTGTCAAATTCCTTGTCTCTGTGCCAGGAATGTGTGGCCACTCTCAATTGACTTCCGTATCGGA | 3877 |
| QY | 1948 | CTGGACAGCCCGAGCGCCCTGGGCCAGCTGCCCCCTGCCTCGAGGTCGAGGGCCAG          | 2007 |
| Db | 3878 | CTAGACAGCCCATCAGAGGGCCCTGGGCCCTGCTGCACCGCTGCAGATGCTGAGGGCCAG      | 3937 |
| QY | 2008 | AAGTGCAGAACTGCTGTGTGCTGAGCGCGGGGAGGGCTACATCGACTTCCGCAATTGGA       | 2067 |
| Db | 3938 | AAGTTGAAGAAATGCACTGGTGTGAGTGGTGAAGTTTACATTGACTTCCGTATCGGA         | 3997 |
| QY | 2068 | GACGAGAGGACGACGAGACGGAGGGCGCAGGGGACATGAGCCAGGTGAAGCCCGTG          | 2127 |
| Db | 3998 | GACGAGAGGATGATGAACCTGAGGAATGTGCCGGGACGTGAACCAAGCAAGCCCTCG         | 4057 |
| QY | 2128 | CTGTCCAAGGCAGCGCAGTCACATCATCGTGTGCGAGGTGTCTTACACCCCGAGTGA         | 2187 |
| Db | 4058 | TTGTCCAAGGCTGAGCGCAGCCACATCATCGTGTGCGAGGTGTCTTACACCCCTGAGTG-      | 4116 |
| QY | 2188 | AGTGTGCTCCCTGCTGGCCCGACCTGTACATAGGACCCCGACCCACCTGACCCCGCC         | 2247 |
| Db | 4117 | AGACCTGTCTCACTGATGCCAACTGTACATAGGACCCCTACCTGCCTGCCTCCCGCCT        | 4176 |
| QY | 2248 | GGCCCGCGGGGTAGCCAGCCAGGGCGCGCCCTCTTCTTAACCTCTCAACCTGCAGCT         | 2307 |
| Db | 4177 | GTTCCCTGGGCGAGCCA-----GGTTCGTCCATCCCTTTTAACCTCTCAACTTGCAGCT       | 4231 |
| QY | 2308 | TTACCTTGAG  | 2317 |
| Db | 4232 | TTTGCCTGAG  | 4241 |

RESULT 10  
AAA62901  
ID AAA62901 standard; DNA; 4269 BP.  
XX  
AC AAA62901;  
XX  
DT 01-NOV-2000 (first entry)  
XX  
DE Murine JNK3 binding protein DNA sequence #3.  
XX  
KW JNK3; Jun N-terminal protein kinase 3; JNK3 binding protein; JSAP;  
KW JNK/SAPK associated protein; nootropic; neuroprotective; neuropleptic;  
KW antiparkinsonian; anticonvulsant; cerebroprotective; antiinflammatory;  
KW nervous system disease; Alzheimer's disease; Parkinson's disease;



Wed Aug 25 09:23:06 2004

|           |   |   |      |
|-----------|---|---|------|
| Db        | 3461  | ACCACAGCACCTGCAGGATGTGGACATAGCCCTATGTTAGCAAGATGCTAGGAACC      | 3520 |
| Qy        | 1588  | GGCAAGCTGGGTTTCTCTTCGTACGCATCAGGCCCCCTGCTTGTGCGGGCAGCCGGCTC   | 1647 |
| Db        | 3521  | GGCAAGCTGGGCTTCTCTTCGTGCGCATCAGCCCTACTCATTCGAGCAACCGTCTG      | 3580 |
| Qy        | 1648  | TGGTGGGACCGGCAACGAGTGGTCTATCTCCATCCCCCTGACAGAGACTGTGTCCTG     | 1707 |
| Db        | 3581  | TGGTGGGCACTGGCAATGGGTTGTCTCTCCATCCCCCTGACTGAGACTGTGTCCTG      | 3640 |
| Qy        | 1708  | CACCGAGGCCAGCTCCTGGGGCTCCGAGCCAATAAGACATCCCCACCTCTGGGGAGGGC   | 1767 |
| Db        | 3641  | CATCGAGGCCAGCTCCTAGGGCTCCGAGGCCAACAAGACATCCCCAACAATCTGGGGAGGG | 3700 |
| Qy        | 1768  | GCCCGTCCCGGGGCAATCATCCAGTGTATGGCGATACACAGTACAGGCGGCCAGC       | 1827 |
| Db        | 3701  | ACCCGCCAGGGGCAATCATCCATGTGTATGGGACGACAGCAGTGACAAGGCCGCAGT     | 3760 |
| Qy        | 1828  | AGCTTCATCCCTACTGCTCCATGGCCCCAGGCCAGCTATGCTTCCATGGGCACCGCGAT   | 1887 |
| Db        | 3761  | AGTTTCATCCCTACTGCTCCATGGCACAGGCTCAGCTTTGCTTCCATGGGCACCGTGAT   | 3820 |
| Qy        | 1888  | GCCGTGAAGTTCTTGTCTCGGTGCCAGGGAACGTGCTGGCCACCCCTGAATGGCAGTGTG  | 1947 |
| Db        | 3821  | GCTGTCAAATTTCTTGTCTGTGCCAGGAATGTGTGGCCACTCTCAATGGCAGTGTG      | 3880 |
| Qy        | 1948  | CTGACAGCCCGAGCGAGGGCCCTGGGCCAGCTGCCCTGCTGGAGGTGCGAGGGCCAG     | 2007 |
| Db        | 3881  | CTAGACAGCCCATCAGAGGGCCCTGGGCCCTGCTGCACCCGCTGCAGATGCTGAGGGCCAG | 3940 |
| Qy        | 2008  | AAGCTGCGGAACGTGCTGGTGGTGGCGGGGAGGCTACATCGACTTCCGCATTGGA       | 2067 |
| Db        | 3941  | AAGTTGAAGAATGCACTGGTGAAGTGGTGAAGTTACATTGACTTCCGTATCGGA        | 4000 |
| Qy        | 2068  | GACGAGAGGACGACGAGACGGAGGGCGCAGGGACATGAGCCAGGTGAAGCCCGTG       | 2127 |
| Db        | 4001  | GACGAGAGGATGATGAAACTGAGGAATGTGCCGGGACGTGAACGACAAAGCCCTCG      | 4060 |
| Qy        | 2128  | CTGTCCAAGGACGAGCGCAGTCACATCATCGTGTGGCAGGTGTCTTACACCCCGAGTGA   | 2187 |
| Db        | 4061  | TTGTCCAAGGCTGAGCGCAGCCACATCATCGTGTGGCAGGTGTCTTACACCCCTGAGTG-  | 4119 |
| Qy        | 2188  | AGTGTGCCCTGCTGGCCCCGACCTGTACATAGGACCCCCGACCACTGACCCCCGCC      | 2247 |
| Db        | 4120  | AGACCTGTCTCTACCTGATGCCAACTGTACATAGGACCTACCTGCTGCTCCCGCCT      | 4179 |
| Qy        | 2248  | GGCCCGGGGTAGCCAGCCAGGGCGCGCCCTTCTTAACCTCTCAACCTGCAGCT         | 2307 |
| Db        | 4180  | GTTCCCTGGGGCAGCCA-----GGTTGCTCCATCCCTTTTAACCTCTCAACTTGCAGCT   | 4234 |
| Qy        | 2308  | TTCACCTGAG  | 2317 |
| Db        | 4235  | TTTGCCCTGAG   | 4244 |
| RESULT 11 |   |   |      |
| ID        | AAS74089  |   |      |
| XX        | AAS74089 standard; cDNA; 2431 BP.   |   |      |
| AC        | AAS74089;   |   |      |
| XX        | 13-FEB-2002 (first entry)   |   |      |
| DT        | DNA encoding novel human diagnostic protein #9893.                        |   |      |
| XX        | Human; chromosome mapping; gene mapping; gene therapy; forensic;          |   |      |
| DE        | food supplement; medical imaging; diagnostic; genetic disorder; ss.       |   |      |
| XX        | Homo sapiens.   |   |      |
| OS        | WO200175067-A2.   |   |      |
| XX        | 11-OCT-2001.  |   |      |
| PN        |   |   |      |
| XX        |   |   |      |
| PD        |   |   |      |
| XX        | 30-MAR-2001;  | 2001WO-US008631.  |      |
| PF        | 31-MAR-2000;  | 2000US-00540217.  |      |
| XX        | 23-AUG-2000;  | 2000US-00649167.  |      |
| PR        | (HYSE-) HYSEQ INC.  |   |      |
| XX        | Drmanac RT, Liu C, Tang YT;   |   |      |
| PA        | WPI; 2001-639362/73.  |   |      |
| PI        | P-PSDB; ABG09902.   |   |      |
| XX        | New isolated polynucleotide and encoded polypeptides, useful in           |   |      |
| DR        | diagnostics, forensics, gene mapping, identification of mutations         |   |      |
| XX        | responsible for genetic disorders or other traits and to assess           |   |      |
| PT        | biodiversity.   |   |      |
| PT        | Claim 1; SEQ ID NO 9893; 103pp; English.                                  |   |      |
| PT        | The invention relates to isolated polynucleotide (I) and polypeptide (II) |   |      |
| XX        | sequences. (I) is useful as hybridisation probes, polymerase chain        |   |      |
| CC        | reaction (PCR) primers, oligomers, and for chromosome and gene mapping,   |   |      |
| CC        | and in recombinant production of (II). The polynucleotides are also used  |   |      |
| CC        | in diagnostics as expressed sequence tags for identifying expressed       |   |      |
| CC        | genes. (I) is useful in gene therapy techniques to restore normal         |   |      |
| CC        | activity of (II) or to treat disease states involving (II). (II) is       |   |      |
| CC        | useful for generating antibodies against it, detecting or quantitating a  |   |      |
| CC        | polypeptide in tissue, as molecular weight markers and as a food          |   |      |
| CC        | supplement. (II) and its binding partners are useful in medical imaging   |   |      |
| CC        | of sites expressing (II). (I) and (II) are useful for treating disorders  |   |      |
| CC        | involving aberrant protein expression or biological activity. The         |   |      |
| CC        | polypeptide and polynucleotide sequences have applications in             |   |      |
| CC        | diagnostics, forensics, gene mapping, identification of mutations         |   |      |
| CC        | responsible for genetic disorders or other traits to assess biodiversity  |   |      |
| CC        | and to produce other types of data and products dependent on DNA and      |   |      |
| CC        | amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  |   |      |
| CC        | coding sequences of the invention. Note: The sequence data for this       |   |      |
| CC        | patent did not appear in the printed specification, but was obtained in   |   |      |
| CC        | electronic format directly from WIPO at                                   |   |      |
| CC        | ftp.wipo.int/pub/published_pct_sequences                                  |   |      |
| XX        | Sequence 2431 BP; 493 A; 764 C; 791 G; 383 T; 0 U; 0 Other;               |   |      |
| SQ        | Query Match 39.5%; Score 1457.4; DB 5; Length 2431;                       |   |      |
|           | Best Local Similarity 93.4%; Pred. No. 7.9e-275;                          |   |      |
|           | Matches 1577; Conservative 0; Mismatches 6; Indels 106; Gaps 1;           |   |      |
| Qy        | 605   | TCTCAGCATCCCGCGCCAGCGACAGCGACTACCTCCCGGGAGATGTTCTCTGGACA      | 664  |
| Db        | 743   | TGTCCTCTCCCGACGCGCCAGCGACAGCGACTACCTCCCGGGAGATGTTCTCTGGACA    | 802  |
| Qy        | 665   | GCGACGTGAACCCAGAGGACCCGGCGCAGATGGCGTCTGCGCGGTATCACCTGTGG      | 724  |
| Db        | 803   | GCGACGTGAACCCAGAGGACCCGGCGCAGATGGCGTCTGCGCGGTATCACCTGTGG      | 862  |
| Qy        | 725   | GCTGTGCCACCCGCTGCAACGTGCCGGGAGCAACTGCTCTCCCGAGGACACCCCGAG     | 784  |
| Db        | 863   | GCTGTGCCACCCGCTGCAACGTGCCGGGAGCAACTGCTCTCCCGAGGACACCCCGAG     | 922  |
| Qy        | 785   | TGCTAGACAAGGGGACGGGGAGGTGGCCACCATCGCCAAACGGGAAGTCAACCCGTCC    | 844  |
| Db        | 923   | TGCTAGACAAGGGGACGGGGAGGTGGCCACCATCGCCAAACGGGAAGTCAACCCGTCC    | 982  |
| Qy        | 845   | AGTCCACAGAGGAGGCCACAGAGGCCACGAGGTGCCAGACCTTGGGCCAGGAGCCAG     | 904  |
| Db        | 983   | AGTCCACAGAGGAGGCCACAGAGGCCACGAGGTGCCAGACCTTGGGCCAGGAGCCAG     | 1042 |
| Qy        | 905   | AGACAGCCACATTGCGGCCCGGGCTCTCACAGAGCACGTTCTCACTGACCCAGCCCCGA   | 964  |
| Db        | 1043  | AGACAGCCACATTGCGGCCCGGGCTCTCACAGAGCACGTTCTCACTGACCCAGCCCCGA   | 1102 |
| Qy        | 965   | CCCCGTCTCTGGCCCCCAGCCTGGCAGCGGAGAACGGGCCAGAGCCTGACAGCAGCAGCA  | 1024 |

Db 1103 CCCCCCTCTCTGGCCCCAGCCTGGCAGCGAGAACGGGCCAGAGCCCTGACAGCAGCAGCA 1162  
QY 1025 CACGGCCAGAGCCAGAGCCAGCGGGGACCCCAACGGGAGCAGGACAGTGTGACACCA 1084  
Db 1163 CACGGCCAGAGCCAGAGCCAGCGGGGACCCCAACGGGAGCAGGACAGTGTGACACCA 1222  
QY 1085 CCATGTGGCTGGGAGCCAGAACGGCTGGCTCTATGTGCACTCGGCTGTGGCCAACTGGA 1144  
Db 1223 CCATGTGGCTGGGAGCCAGAACGGCTGGCTCTATGTGCACTCGGCTGTGGCCAACTGGA 1282  
QY 1145 AGAAGTGCCTGCACCTCCATCAAGCTGAAGGATCTGTGCTGAGCCTGGTGTGCAAAAG 1204  
Db 1283 AGAAGTGCCTGCACCTCCATCAAGCTGAAGGATCTGTGCTGAGCCTGGTGTGCAAAAG 1342  
QY 1205 GCGGTGTGCTGGTGGCTCTGGCGGACGGGACCCCTGGCCATCTTCCACCGTGGTGAAGATG 1264  
Db 1343 GCGGTGTGCTGGTGGCTCTGGCGGACGGGACCCCTGGCCATCTTCCACCGTGGTGAAGATG 1402  
QY 1265 GCCAGTGGGATCTGAGCAACTATCACCTAATGGACCTGGGCCACCCGACCCACTCCATCC 1324  
Db 1403 GCCAGTGGGATCTGAGCAACTATCACCTAATGGACCTGGGCCACCCGACCCACTCCATCC 1462  
QY 1325 GCTGCATGGCTGTGTGTACGACCGCGTGTGTGGCTACAGAAACAAGGTGCACGTCA 1384  
Db 1463 GCTGCATGGCTGTGTGTACGACCGCGTGTGTGGCTACAGAAACAAGGTGCACGTCA 1522  
QY 1385 TCCAGCCCCAAGACCATGCAGATAG----- 1408  
Db 1523 TCCAGCCCCAAGACCATGCAGATAGAGCGGAGTGCAGGCCCGGGAGGGAAGA 1582  
QY 1409 ----- 1408  
Db 1583 GGCTCCTGCTGGCCAGCAGCTCTCCGCACTCTCCATACCGGAAGTCCACGAGGCTTACCC 1642  
QY 1409 -----AGAAAGTCAATTGACGCCCCACCGCGGGAGAGCCAGGTGCGGCAGCTG 1458  
Db 1643 ATGACTCCAGAAAGTCAATTGACGCCCCACCGCGGGGAGAGCCAGGTGCGGCAGCTG 1702  
QY 1459 GCGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCTGGACTCCACCCCTGAGGCTCTAC 1518  
Db 1703 GCGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCTGGACTCCACCCCTGAGGCTCTAC 1762  
QY 1519 CATGCACACACGCCAGCATCTACAGGACGTGGACATTCAGGACCCCTACGTCAGCAAGATG 1578  
Db 1763 CATGCACACACGCCAGCATCTACAGGACGTGGACATTCAGGACCCCTACGTCAGCAAGATG 1822  
QY 1579 CTAGGCACTGGCAAGCTGGGTTTCTCTCGTACGCATCACGGCCCTGCTGTGCGGGGC 1638  
Db 1823 CTAGGCACTGGCAAGCTGGGTTTCTCTCGTACGCATCACGGCCCTGCTGTGCGGGGC 1882  
QY 1639 AGCCGGCTCTGGTGGGACCGGCAACGAGTGGTCACTCCATCCCTGACAGAGACT 1698  
Db 1883 AGCCGGCTCTGGTGGGACCGGCAACGAGTGGTCACTCCATCCCTGACAGAGACT 1942  
QY 1699 GTGGTCTGCACCGAGGCGAGCTCCTGGGGCTCCGAGGCAATAAGACATCCCCACCTCT 1758  
Db 1943 GTGGTCTGCACCGAGGCGAGCTCCTGGGGCTCCGAGGCAATAAGACATCCCCACCTCT 2002  
QY 1759 GGGAGGGCGCCCGTCCCGGGGCACTCATCCAGTGTATGGCGATGACAGCAGTGACAGG 1818  
Db 2003 GGGAGGGCGCCCGTCCCGGGGCACTCATCCAGTGTATGGCGATGACAGCAGTGACAGG 2062  
QY 1819 GCGGCAGCAGCTTCACTCCCTACTGCTCCATGGCCCGGCGGAGTGTGCTTCCATGGG 1878  
Db 2063 GCGGCAGCAGCTTCACTCCCTACTGCTCCATGGCCCGGCGGAGTGTGCTTCCATGGG 2122  
QY 1879 CACCGGATGCCGTGAAGTTCTTTGCTCGGTGCCAGGGAACGTGTGGCCACCCCTGAAT 1938  
Db 2123 CACCGGATGCCGTGAAGTTCTTTGCTCGGTGCCAGGGAACGTGTGGCCACCCCTGAAT 2182  
QY 1939 GGCAGTGTGTCAGACCCAGCCGAGGCGGCTGGCCAGTGGCCCTGCTCGGAGGTC 1998

Db 2183 GGCACTGTGTGGACACCCAGCCAGCGGGCCCTGGGCCAGTGCCTCGGAGGTC 2242  
QY 1999 GAGGGCCAGAAGCTGCGGAACGTGCTGGTGTGCTGAGCGGGGGAGGGTACATCGACTTC 2058  
Db 2243 GAGGGCCAGAAGCTGCGGAACGTGCTGGTGTGCTGAGCGGGGGAGGGTACATCGACTTC 2302  
QY 2059 CGCATTTGGAGACGGAGAGGACGACGAGACGGAGGAGGCGCAGGGGACATGAGCCAGTG 2118  
Db 2303 CGCATTTGGAGACGGAGAGGACGACGAGACGGAGGAGGCGCAGGGGACATGAGCCAGTG 2362  
QY 2119 AAGCCCGTGTGTCCAAAGGCGAGGCGCAGTCACATCATCTGTGTGGAGGTGTCTACACC 2178  
Db 2363 AAGCCCGTGTGTCCAAAGGCGAGGCGCAGTCACATCATCTGTGTGGAGGTGTCTACACC 2422  
QY 2179 CCCGAGTGA 2187  
Db 2423 CCCGAGTGA 2431  
RESULT 12  
ABA09028  
ID ABA09028 standard; cDNA; 4688 BP.  
XX ABA09028;  
AC ABA09028;  
XX  
DT 11-JAN-2002 (first entry)  
XX  
DE Human secreted protein homologue-encoding cDNA, SEQ ID NO:804.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;  
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
KW inhibitor; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
KW chronic inflammatory condition; proliferative retinopathy;  
KW atherosclerosis; coronary heart disease; arterial ischaemia;  
KW bone disorder; osteoporosis; vascular growth disorder;  
KW tissue regeneration; wound healing; infection; immune disorder;  
KW cell culture; drug screening; gene therapy; antiinflammatory;  
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
KW antifungal; vulnerary; antiulcer; ss.  
XX Homo sapiens.  
OS  
XX  
XX WO200157188-A2.  
PN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 05-FEB-2001; 2001WO-US003800.  
PF  
XX  
XX 03-FEB-2000; 2000US-00496914.  
PR  
XX 27-APR-2000; 2000US-00560875.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
PI  
XX  
DR WPI; 2001-457740/49.  
DR P-PSDB; ABB11784.  
XX  
PT Human proteins and DNA encoding sequences useful for preventing, treating  
PT or ameliorating a medical condition in a mammalian subject e.g. arthritis  
PT and cancer.  
XX  
PS Claim 1; Page 710-712; 1963pp; English.  
XX  
CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
CC invention also relates to vectors and recombinant host cells comprising a  
CC nucleotide of the invention, methods of producing the novel polypeptides,  
CC antibodies against the polypeptides, methods of detecting the nucleotides  
CC or polypeptides in a sample, and methods of identifying compounds which

CC bind to polypeptides of the invention. Although novel, many of the  
CC polypeptides of the invention have homology to known proteins, thereby  
CC giving an insight into their probable biological activities, and hence  
CC potential therapeutic applications. The polypeptides of the invention may  
CC have various activities, including cytokine, cell proliferation or cell  
CC differentiation activities; stem cell growth factor activity;  
CC haematopoiesis regulatory activity; tissue growth factor activity;  
CC immunomodulatory activity; activin- or inhibin-related activities;  
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
CC thrombolytic activities; receptor or ligand activities; or may be  
CC involved in oncogenesis, cancer cell proliferation or metastasis.  
CC The invention on their biological activities, polypeptides and nucleotides of  
CC the invention are useful for preventing, treating or ameliorating medical  
CC conditions, e.g., by protein or gene therapy. Such conditions include  
CC cancers, haematopoietic disorders (e.g. myeloid or lymphoid cell  
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
CC vascular growth. Polypeptides involved with tissue regeneration and  
CC repair (or nucleic acids encoding them) may be used to promote wound  
CC healing (e.g., of burns, incisions and ulcers), while those with  
CC immunomodulatory activities may be used in the treatment of viral,  
CC bacterial and fungal infections in addition to immune disorders.  
CC Polypeptides with growth factor activity may be used in cell cultures to  
CC promote cell growth. For example, such polypeptides may be used to  
CC manipulate stem cells in culture to give rise to neuroepithelial cells  
CC that can be used to augment or replace cells damaged by illness,  
CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
CC may also be used in the diagnosis of the above conditions, and in drug  
CC screening techniques. The present sequence represents a cDNA encoding a  
CC novel human polypeptide of the invention  
XX

SQ Sequence 4688 BP; 1438 A; 931 C; 1163 G; 1156 T; 0 U; 0 Other;

|                       |  |                  |            |              |
|-----------------------|--|------------------|------------|--------------|
| Query Match           | 13.2%;   | Score 486.2;     | DB 4;      | Length 4688; |
| Best Local Similarity | 58.5%;   | Pred. No. 3e-85; |            |              |
| Matches 1191;         | Conservative 0;  | Mismatches 753;  | Indels 93; | Gaps 16;     |
| QY 214                | GGCCAGGAGGACACGCGGATGAAGAACGTGCGGTGCGGTGCTACTGCGCCCTCTGGTG     | 273              |            |              |
| Db 1973               | GGTCAAGGTGAAATAAGATGAAAAATTACCTGTGCTGTCTATCTCAGACCTCTGGAT      | 2032             |            |              |
| QY 274                | GAGAAAGACCCACCATGAAGCTGTGTGTGCGCGCGCGTCAACCTGAGCGGGTGGAGG      | 333              |            |              |
| Db 2033               | GAAAAAGATACATCAATGAAGCTGTGTGTGCTGTTGGAGTCAATTTATCTGTGGGAAG     | 2092             |            |              |
| QY 334                | CCCAATGAGGACGACGCTGGGAATGGAG-----TCAAGCCAGCGCCAGCGCGC          | 381              |            |              |
| Db 2093               | ACCAGAGATGGTGGTTCTGTTGTTGGCAAGTGTATTTACAAGGATGTTGCTGTTG        | 2152             |            |              |
| QY 382                | GATCCCTGTACCTGCGACCGGAGGAGACGGCGGAGCCCAAGAGCGGCCACACGCTCCCC    | 441              |            |              |
| Db 2153               | GATACAGAAGGCAGTAACAGCGAAGTGCCTCTCAGAGTAGTTTAGATAAGTTAGATCAG    | 2212             |            |              |
| QY 442                | GAG---AAGAAGAAGGCCAAGGAGCTCCCTGAAATGGAGCGCCACCTCCAGCCGGGTGCG   | 498              |            |              |
| Db 2213               | GAACTTAAGGAACAGCAGAAGGAGTTAAAAAATCAAGAAGAATTATCCAGTCTAGTTGG    | 2272             |            |              |
| QY 499                | ATCCTGACAGCACCCCTGACCAACAGCAAGGTGGTGTATCATCGACGCCCAACAGCCGGGC  | 558              |            |              |
| Db 2273               | ATCTGTACCAGCACTCAITCGGCTACAAAAGTTCTTATTTATTTAGTGTCTTCAACCTGGC  | 2332             |            |              |
| QY 559                | ACGGTGGTGACCAAGTTACCGTCTGCAACCGGCACGTTGTTGTCATCTCCAGCATCCCC    | 618              |            |              |
| Db 2333               | AACATCTTAGACAGTTTCACTGTTTGGCAACTCTCATGTTCTGTGCAATTGCAAGTGTGCCA | 2392             |            |              |
| QY 619                | GCGGCCAGGCACAGCGACTACCTCCCGG--GGAGATGTTCTCTGGACAGCGACGTGAACCC  | 677              |            |              |
| Db 2393               | GGTGACAGAAACAGACTACCTGCGAGGAGAAGATCTTTCAGAAATCTGTCAGGTAGAC     | 2452             |            |              |
| QY 678                | AGAGGACC-----CGGGCGCAGATGGCGTGTG                               | 705              |            |              |
| Db 2453               | AAAGCATCTTTATGTGGAAGTATGACAAAGCATCAGCAGACAGACAGACAGCTGTGA      | 2512             |            |              |

|         |  |      |  |  |
|---------|--|------|--|--|
| QY 706  | GCCGGTATCACCCCTGGTGGGTGTGTC--CACCCGCTGCAACGTCGCCGCGAGCAACTGCT  | 763  |  |  |
| Db 2513 | GGAGGCATCACAGTGGTGGTGTGTTCTGCAGAAGGTGTACGGGAGTGCACACTTCCCT     | 2572 |  |  |
| QY 764  | CCTCCCGAGGGGACACC--CCAGTGTCTAGACAAAGGGGACGGGGAGGTGGCCACCATCGCC | 822  |  |  |
| Db 2573 | AGTACAAATGGTGTCTTCTCCAGTGATGGATAAACACCCAGAAATGGAGCAGAAATAGT    | 2632 |  |  |
| QY 823  | AACGGGAAGGTCAACCCGTCCTCCAGTCCACAGAGGAGGCCACAGAGGCCACGGAGGTGCCA | 882  |  |  |
| Db 2633 | GA--GGTTGATGAAATGTTCACACAGCAGAAAGAACTGAAGCTACAGAAGG--A         | 2688 |  |  |
| QY 883  | GACCTGGGCCAGCGAGCCAGAGACAGCCACATTTGGGCCCGGGCCTCTCACAGAGCAC     | 942  |  |  |
| Db 2689 | ATCGGGGTGAGCTGAAGACACAGTTGGACATCTCCCAAATGGCGTCTACACAGAGCAT     | 2748 |  |  |
| QY 943  | GTCTTCACTGACCCAG-----CCCCGACCCCGTCTCTGGCCCCCAGCC               | 986  |  |  |
| Db 2749 | GTCTTTACAGATCCTTTGGGAAGTTCAGATCCAGAGACCTCTCCCAGTGTATCAGTC      | 2808 |  |  |
| QY 987  | TGGCAGCGAGAACGGGCCAGAGCCTGACAGCAGCAGCACACGGCCA---GAGCCAGAGCC   | 1043 |  |  |
| Db 2809 | GAGCAATGACTCAGATGCATATAAAGATCAAAATATCAGTACTGCCAAATGAACAAGACTT  | 2868 |  |  |
| QY 1044 | CAGCGGGGACCCACGGGAGCAGGCGAGCTGCTGCACCCACATGTGGCTGGGAGCCCA      | 1103 |  |  |
| Db 2869 | GCTGAGAGAAGAAGCCAGAAAATGAGTAGTCTTTTACCACTATGTGGCTGGAGCTCA      | 2928 |  |  |
| QY 1104 | GAACGGCTGGCTCTATGTGCTACTCGGCTGTGGCCAACTGGAAGAAGTGCCTGCACTCCAT  | 1163 |  |  |
| Db 2929 | AAATGGCTGTTGTATGTCCATTCATCTGTAGCCAGTGGAGAAATGTCTCCATTCAT       | 2988 |  |  |
| QY 1164 | CAAGCTGAAGGATTTCTGTGCTGAGCCTGGTGCAATGTCAAAGGCCGTGTGCTGGTGTCT   | 1223 |  |  |
| Db 2989 | TAAACTTAAAGATTCGATTTCTCAGTATTGTACAGTGAAGGGAATCGTGTGTAGTACCT    | 3048 |  |  |
| QY 1224 | GGCGGACGGGACCCCTGGCCATCTTCCACCGTGTGAAGATGGCCAGTGGGATCTGAGCAA   | 1283 |  |  |
| Db 3049 | GGCTGACGGCACCCCTTGCAATCTTTCACAGAGGAGTGGATGGGAGTGGGATTTGTCAA    | 3108 |  |  |
| QY 1284 | CTATCACCTAATGACCTGGGCCACCCGACCACTCCATCCGCTGCATGGCTGTGTGTATA    | 1343 |  |  |
| Db 3109 | CTATCACCTCTTAGACCTTGGACGGCCTCATCATTCATCCGTTGCAATGACTGTGGTACA   | 3168 |  |  |
| QY 1344 | CGACCGCGTGTGTGGTGTACAGAAACAAGGTGCACGTCAATCCAGCCCAAGACCATGCA    | 1403 |  |  |
| Db 3169 | TGACAAAAGTCTGGTGGCTATAGGAAACAAAATCTATGTGTGTCAGCCAAAGGCCATGA    | 3228 |  |  |
| QY 1404 | GATAGAGAAAGTCATTTGACGCCCCACCGCGGGGAGAGCCAGGTGCGGCAGCTGGCGTG    | 1463 |  |  |
| Db 3229 | AATAGAGAAATCTTTTGATGCACATCCAGGAAGGAGAGCCAGTGCACAGCTTGCCTG      | 3288 |  |  |
| QY 1464 | GATCGCGGATGGGTATGGTGTCCATCCGCTTGGACTCCACCTGAGGCTCTACCATGC      | 1523 |  |  |
| Db 3289 | GGTGGGGATGGCGTGGGTCTCCATTCGTTGGATTCTACGCTCCGTCTCTATCATGC       | 3348 |  |  |
| QY 1524 | ACACAGCACCCAGCATCTACAGGACGTGGACATTTAGCCCTACGTACGCAAGATGCTAGG   | 1583 |  |  |
| Db 3349 | ACACACTTATCAACATCTACAGGATGTGGACATTTAGCCTTATTAAGCAAAATGTTAGG    | 3408 |  |  |
| QY 1584 | CACCTGCAAGCTGGTGTCTCTTCGTACGCATCAAGGCCCTGTCTGTGCGGGCAGCGG      | 1643 |  |  |
| Db 3409 | TACTGGAATACTGGCTTCTCTTTTGTGAGAATTACAGCTCTTATGGTGTCTTGTAAATCG   | 3468 |  |  |
| QY 1644 | GCTCTGGGTGGGCACCGGCAACGGAGTGGTCTCTCCATCCCTCCCTGACAGAGACTGTGGT  | 1703 |  |  |
| Db 3469 | TTTGTGGGTGGGACAGGAAATGGTGTCTATTTCTCCATCCCATTTGACAGAAACCGTAAT   | 3528 |  |  |
| QY 1704 | CCTGCACCGAGGCCAGCTCTCTGGGCTCCGAGCCCAATAGACATCCCCACCTCTGGGGA    | 1763 |  |  |
| Db 3529 | CCTCCACCGAGGACGTTTACTGGGGCTGAGGGCAATAAACCTCAGGTGTACCAGGAAA     | 3588 |  |  |



Db 1630 AGTACAAATGGTGCTTCTCCAGTGATGGATAAAACCACAGAAATGGAAGCAGAAAAATAGT 1689

Qy 826 GGGAAGGTCAACCCGTCAGTCCAGTCCACAGAGGAGGCCACAGAGGCCACCGAGGTGCCAGAC 885

Db 1690 GAGGTTGATGAAATGTTCCAAACAGCAGAGAAGCAACTGAAGCTACAGAAGGA---AT 1746

Qy 886 CCTGGGCCAGCGAGCCAGAGACAGCCACATTCGGGCCCGGCTCTCTCACAGAGCACGTC 945

Db 1747 GCGGGGTAGCTGAAGACACAGTGGACATCTCCCAAACCTGCGCTCTACACAGAGCATGTC 1806

Qy 946 TTCACTGACCC-----AGCCCCGACCCCGTCTCTGTGGCCCCCAGCCTGGC 990

Db 1807 TTACAGATCCTTTGGGAGTTTCAGATCCCAGAAGACCTCTCCCAGTGTATCAGTCGAGC 1866

Qy 991 AGCGAGAACGGGCCAGAGCCTTGACAGCAGCAGCACAGGCCA---GAGCCAGAGCCCCAGC 1047

Db 1867 AATGACTCAGATGCATATAAAGATCAAAATATCAGTACTGCCAAATGAACAAGACTTGGTG 1926

Qy 1048 GGGGACCCACGGGAGCAGGCAGCAGTGTGCAACCCACCATGTGGCTGGGAGCCAGAAC 1107

Db 1927 AGAGAAGAAGCCAGAAATGAGTAGTCTTTTACCAACTATGTGGCTTGGAGCTCAAAAT 1986

Qy 1108 GGTGCTCTATGTGCACTCGGCTGTGGCCAACTGGAAGAGTGCCTGCACTCCATCAAG 1167

Db 1987 GGCTGTTGTATGTCATTCTCAGTATTGTACACGTGAAGGAATCGTGTAGTAGCCCTGGCT 2046

Qy 1168 CTGAAGGATTCTGTGCTGAGCCTGTGTCATGTCAAAAGCCGTGTGCTGGTCTGGCG 1227

Db 2047 CTTAAAGATTGATTCGATTCTCAGTATTGTACACGTGAAGGAATCGTGTAGTAGCCCTGGCT 2106

Qy 1228 GACGGGACCCCTGGCCATCTTCCACCGTGGTGAAGATGGCAGTGGGATCTGAGCAACTAT 1287

Db 2107 GACGGCACCCCTTGCAATCTTTCACAGAGGAGTGGATGGGAGTGGGATTTGTCAAACTAT 2166

Qy 1288 CACCTAATGGAACCTGGGCCACCCGCAACCACTCCATCCGCTGCATGGCTGTTGTACGAC 1347

Db 2167 CACCTCTTAGACCTTGGACGGCTCATATTCATCCGTTGATGACTGTGGTACATGAC 2226

Qy 1348 CGCGTGTGGTGTGGCTACAAGAACAAAGGTGCACGTTCATCCAGCCCCAAGACCATGCAGATA 1407

Db 2227 AAAGTCTGGTGTGGCTATAGGAACAAAATCTATGTGTGCAGCCAAAGGCCATGAAAATA 2286

Qy 1408 GAGAAATGATTTGACGCCACCCGCGGGAGAGCCAGGTGCGGAGCTGGCGTGGATC 1467

Db 2287 GAGAAATCTTTTGATGCACATCCCAGGAAGAGAGCCAAAGTGGACAGCTTGGTGGTG 2346

Qy 1468 GCGATGGCGTATGGGTGTCCATCCGCTGGACTCCACCCCTGAGGCTCTACCATGCAAC 1527

Db 2347 GGGGATGGCGTGGGTCTCCATTGCTTGGATTCTACGCTCCGTTCTCTATCATGCAAC 2406

Qy 1528 ACGCACAGCATCTACAGGACGTGGACATGAGCCCTACGTACAGCAAGATGCTAGGCACT 1587

Db 2407 ACTTATCAACATCTACAGATGTGGACATGAGCCTTATGTAAGCAAAATGTTAGGTACT 2466

Qy 1588 GGCAAGCTGGGTTTCTCCTTCGTACGCATCACGGCCCTGCTTGTGCGGGCAGCCGGCTC 1647

Db 2467 GGAAACCTGGGCTTCTCTTTGTGAGAAATACAGCTCTATGGTGTCTTGTAAATCGTTG 2526

Qy 1648 TGGGTGGGACCGGCAACGGAGTGGTTCATCTCCATCCCTTGACAGAGACTGTGGTCCTG 1707

Db 2527 TGGGTGGGACAGGAAATGGTGTCTATTATCTCCATCCCAATTGACAGAAACCGTAATCCTC 2586

Qy 1708 CACCGAGGCAGCTCCTGGGCTCCGAGCCCAATAAGACATCCCCACCTCTGGGGAGGGC 1767

Db 2587 CACCAGGACGTTTACTGGGGCTGAGGGCAATAAAACCTCAGGTGTACCAGGAAAT--- 2643

Qy 1768 GCCCGTCCGGGGCATCATCCACGTGTATGGGATGACAGCAGTGAACA---GGGCGGCC 1824

Db 2644 ---CGTCTGGAAGTGTAATCCGTGTATATGGTGTATGATAAAGAGTGATAAAGTGACTCCA 2700

Qy 1825 AGCAGCTTCATCCCTACTGCTCCATGGCCAGCCAGCCAGCTATGCTTCCATGGGCACCGC 1884

Db 2701 GGGACATTTTATACCCCTATTGTTCAATGGCACATGCACAGCTTTGCTTCCATGGGCACCGG 2760

Qy 1885 GATGCCGTGAAGTTCTTTGTCTCGTGCCAGGAACTGCTGGCCACCTGAATGGCAGT 1944

Db 2761 GATGCTGTGAAATCTTTGTGGCAGTCCAGGTCAAGTCATCAGCCCAAAAAGTAGCAGT 2820

Qy 1945 GTGCTGGACAGCCCAGCCAGGGGCCCTGGGCCAGCTGCCCTCGAGGTCGAGGGC 2004

Db 2821 AGTGGCACGGATCTGACGGGTGACAAAGCAGGGCCATCTGCACAGGAGCCTGGTAGTCAG 2880

Qy 2005 CAGAAAGTTCGGGAACGTGCTGGTGTGCTGAGCGGGCGGGAGGGCTACATCGACTTCCGCAAT 2064

Db 2881 ACGCCCTTGAAGTCTATGCTTGTTCATCAGTGGAGGAGGGGTACATCGACTTCCGAATG 2940

Qy 2065 GGAGACGGAGAGGACGACGAGACGGAGAGGGCGCAGGGGACATGAGCCAGGTGAAGCCC 2124

Db 2941 GGTGATGAAG-----GTGGAGAAATCAGAACTTCTTTGGAGAGGATCTTCCACTTGAACCTT 2995

Qy 2125 GTGCTGTCCAAGGCAGAGCGCAGTCACATCATGCTGTGGCAGGTGTCTTACACCCCGCAG 2184

Db 2996 CTG-TCACCAAAGCAGAAAGGAGTCACTTGATAGTGTGGCAAGTGATGTATGGCAATGAG 3054

Qy 2185 TGA 2187

Db 3055 TGA 3057

RESULT 14

ADA45191

ID ADA45191 standard; cDNA; 4667 BP.

XX

AC ADA45191;

XX

DT 20-NOV-2003 (first entry)

XX

DE Murine JLP coding sequence.

XX

KW Cytostatic; Gene therapy; murine; scaffolding protein; JLP;

KW JNK-associated Leucine zipper Protein; MEK kinase 3; MEKK3;

KW MAP kinase kinase 4; MKK4; c-Jun NH2-terminal kinase; JNK;

KW p38 MAP kinase; MAPK; c-Myc; MAX; apoptosis; cancer; gene; ss.

XX

OS Mus musculus.

XX

FH Key

FT CDS

FT Location/Qualifiers

FT 163..4086

FT /\*tag= a

FT /product= "Murine JLP"

WO2003066652-A2.

14-AUG-2003.

04-FEB-2003; 2003WO-US003355.

05-FEB-2002; 2002US-0354377P.

(UTEM ) UNIV TEMPLE.

Lee CM, Dhanasekaran N, Reddy PE;

WPI; 2003-731487/69.

P-PSDB; ADA45192.

New scaffolding nucleic acid sequences, designated as JLP, useful for modulating apoptotic response in a cell, and thus for treating metastatic cancer.

Claim 10; Page 79-80; 102pp; English.

The present invention relates to novel human and murine scaffolding proteins, JLP (for JNK-associated Leucine zipper Protein, ADA45190 and ADA45192). JLP tethers MEK Kinase 3 (MEKK3), Mitogen-Activated Protein (MAP) kinase kinase 4 (MKK4), c-Jun NH2-terminal kinase (JNK), p38 MAP

CC kinase (MAPK), c-Myc and MAX into a signalling module which controls the  
CC apoptotic response. JLP therefore functions as a signalling conduit to  
CC transmit extracellular signals to the nucleus through MEK3-MKK4-  
CC JNK/p38/MAPK/c-Myc/MAX signalling module. The JLP sequences are useful  
CC for modulating apoptotic response in a cell, and thus for treating  
CC metastatic cancer. The present sequence is the coding sequence for murine  
CC JLP.  
XX  
SQ Sequence 4667 BP; 1376 A; 1024 C; 1201 G; 1066 T; 0 U; 0 Other;

Query Match 12.0%; Score 442.2; DB 8; Length 4667;  
Best Local Similarity 56.5%; Pred. No. 1.1e-76;  
Matches 1159; Conservative 0; Mismatches 763; Indels 129; Gaps 13;

|    |      |  |      |
|----|------|--|------|
| QY | 214  | GGCCAGGAGGACACGCGGATGAAGAACGTGCGGTGCGCGGCGTCAACCTGAGCGGGTGGAGG | 273  |
| Db | 2098 | GGTCAAGGGGAACTAAGATGAAAAATTACCTGTGCTGTCTATCTCAGACCTAGAC        | 2157 |
| QY | 274  | GAGAGGACCCACCATGAAGCTGTGGTGTGCGCGGCGTCAACCTGAGCGGGTGGAGG       | 333  |
| Db | 2158 | GAAAAAGATGCGTCTATGAAGCTGTGGTGTGCTGTGAGGAGTCAATTATCTGGTGCAAG    | 2217 |
| QY | 334  | CCCAATGAGGACGACGCTGGGAATGGAG-----TCAGCCAGCGCCAGCGCGC           | 381  |
| Db | 2218 | ACAAGAGATGGCGGTCTGTTGTGGAGCGAGTGTATTTTACAGGATATTGCTGGTTG       | 2277 |
| QY | 382  | GATCCCCCTGACCTGCGACCGCGAAGGAGACGGCGAGCCCAAGAGCGGCCACAGCTCTCCC  | 441  |
| Db | 2278 | GACACTGAAGCGAGTAACACAGCGAGTGCGTCTCAGAGTAGTTTAGACAAGTTGGATCAA   | 2337 |
| QY | 442  | GA---GAAGAAAGAGGCCAAGGAGCTCCCTGAAATGGAAGCCACCTCCAGCGGGTGTGG    | 498  |
| Db | 2338 | GAACTCAAGGAACAGCAGAAAGAAATTAAATCAAGAGAGTTGTCAGTCTGG            | 2397 |
| QY | 499  | ATCCTGACCCAGCACCTGACCAAGCAGCAAGGTGGTGTATCATCAGAGCCCAACAGCGGGC  | 558  |
| Db | 2398 | ATCTGTACCAAGCACCCACTCAACTACAAGGTGATCATCATGATGCTGTTTCAAGCCTGGC  | 2457 |
| QY | 559  | ACGGTGGTGACCAAGTTACCGTCTGCAACGGCACGTGCTGTGTCATCTCCAGCATCCCC    | 618  |
| Db | 2458 | AACATCCTAGATAGTTTCACTGTTTGCAATTCTCATGTTCTGTGTCATGGCCAGTGTCCCA  | 2517 |
| QY | 619  | GCGGCCAGGACAGCACTACCTCCCGG-GGAGATGTTCTCTGGACAGCGACGTGAACCC     | 677  |
| Db | 2518 | GGAGCTCGGAAACAGACTACCTGTCACAGAGAGAACTTCTGAATCTGGTCTAGGTAGAC    | 2577 |
| QY | 678  | AGAGGACCGGGGCGCAGATG-----GCGTGTGTG                             | 705  |
| Db | 2578 | AAAGCGTCATTGTGTGAAGCATGACCCAGTAACAGCTCCGACAGATGGACAGCTTGCTG    | 2637 |
| QY | 706  | GCCGGTATCACCTGTGGGTGTGCCACCGCTGCAACGTGCCGGAGCAACTGCTCC         | 765  |
| Db | 2638 | GGAGGCATCACAGTGGTGGTGTCTCCACAGAGGACTGACAGGAGTGCCTCCCA          | 2697 |
| QY | 766  | TCCGAGGGGACACCCAGTGTAGACAAGGGGACGGGGAGGTGGCCACCATCGCCAAC       | 825  |
| Db | 2698 | AGTACCAACGGTGTCTCTGTGATTGAAAGCCACAGAAATGGAAATAGC               | 2757 |
| QY | 826  | GGGAAGGTCAACCCGTCCTCCAGTCCACAGAGGAGCCACAGAGGCCACGGAGGTGCCAGAC  | 885  |
| Db | 2758 | GAAGTGTGATGAAATATTCCAACAGCAGAGAGCACTGAAGCCACAGAGG---GCAAT      | 2814 |
| QY | 886  | CCTGGGCCAGCGAGCCAGACAGCCACATTGGCGCCCGGCGCTCTCACAGAGCAGTTC      | 945  |
| Db | 2815 | GCAGGTCCACTGAAGACACTGTGGACATCTCCAGCTGGCGTGTACACAGAGCATGTG      | 2874 |
| QY | 946  | TTCATGACCCA-----GCCCGACCCCGTCTCTGGCCCGCCAGCTGGCAGCGAGAA        | 998  |
| Db | 2875 | TTTACAGATCCACTGGGAGTTCAGATCCACAGAGACCTCTCCCGAGTGTTCAGTCAAGT    | 2934 |
| QY | 999  | CGGGCCAGAGCCTGACAGCAGCAGACA-----CGGCCAGAGCCAGAGCCAGC           | 1047 |
| Db | 2935 | AATGACTCAGATGTGTACAAAGATCAGATATCAGTATTGCCAATGAACAAGACCTTGG     | 2994 |

|    |      |  |      |
|----|------|--|------|
| QY | 1048 | GGGGACCCACGGGAGCAGGCAGCTGCTGCACCCACCATGTGGCTGGAGCCAGAAC        | 1107 |
| Db | 2995 | AGAGAGAAGCTCAGAAAATGAGTAGTCTTTTACCACCATGTGGCTTGAGCTCAGAA       | 3054 |
| QY | 1108 | GGCTGGCTCTATGTGCACTCGGCTGTGGCCAACTGGAGAAAGTGCCTGCACCTCCATCAAG  | 1167 |
| Db | 3055 | GGCTGTTTGTATGTCCTCATCTGTAGCCAGTGGAGGAAATGTCTCCATTCATTAAG       | 3114 |
| QY | 1168 | CTGAAGGATTCTGTGCTGAGCCTGGTGCATGTCAAGGCCGTGTGTGGTGGCTCTGGCG     | 1227 |
| Db | 3115 | CTCAAGACTCAATACTCAGTATTGTACATGTGAAGGAATTTGTACTAGTGGCCCTGGCG    | 3174 |
| QY | 1228 | GACGGACCCCTGGCCATCTTCCACCGTGGTGAAGATGGCCAGTGGGATCTGAGCACTAT    | 1287 |
| Db | 3175 | GATGGCACCCCTTGCATCTTCCACAGAGGAGTTGATGGACAATGGGATCTGTCAAACTAT   | 3234 |
| QY | 1288 | CACCTAATGGACCTGGGCCACCCGCAACCATCCATCCGCTGCATGGCTGTTGTGTACGAC   | 1347 |
| Db | 3235 | CACCTTTTAGACCTTGGACGCTCCTCATCTCCATACGATGCATGACTGTGTACATGAC     | 3294 |
| QY | 1348 | CGCGTGTGTGGCTACAAGAACAAAGGTGCACTCATCCAGCCCAAGACCATGCAGATA      | 1407 |
| Db | 3295 | AAAGTCTGGTGGCTATAGGAACAAAATCTATGTGTTTCAACCAAGGCTATGAAGATA      | 3354 |
| QY | 1408 | GAGAACTCATTTGACGCCCCACCGCGGGGAGAGCCAGGTGGCGGAGCTGGCGTGGATC     | 1467 |
| Db | 3355 | GAGAAATCATTTGATGCACACCCAGGAAGAGAGCCAAAGTACGGCAGCTTGCACTGGTG    | 3414 |
| QY | 1468 | GGCGATGGCGTATGGGTGTCCATCCGCTGGACTCCACCTGAGGCTCTACCATGCACAC     | 1527 |
| Db | 3415 | GGTGTGGTGTGGGTCTCCATTCGTTGGATTCCACGCTCCGCTCTATCATGCACAT        | 3474 |
| QY | 1528 | ACGCACCAAGCATCTACAGGAGCTGGACATTTGAGCCCTACGTCAGCAAGATGCTAGGCACT | 1587 |
| Db | 3475 | ACATACCAACATCTCCAGGATGTGGATATTGAGCCCTTATGTAAAGCAAAATGTAGGTACT  | 3534 |
| QY | 1588 | GGCAAGCTGGGTTTCTCCTTCGTACGCATCACGGCCCTGCTTGTGCGGGGAGCCGGCTC    | 1647 |
| Db | 3535 | GGAAAACCTGGGTTTCTCCTTCGTGAGGATAAACAGCCCTTATGTTGCTTTCGCAATCGTTG | 3594 |
| QY | 1648 | TGGGTGGGACCGGCAACGGAGTGGTCACTTCCATCCCTGACAGAGACTGTGGTCTCTG     | 1707 |
| Db | 3595 | TGGGTGGGACTGGAATGGTGTCACTTATCTCCATCCCACTGACAGAAACA-----        | 3645 |
| QY | 1708 | CACCGAGGCCAGCTCCTGGGGCTCCGAGCCAATAAGACATCCCCCACCTTGGGGAGGGC    | 1767 |
| Db | 3646 | -----AATAAACTCAGGACTCCAGTTAA-----                              | 3671 |
| QY | 1768 | GCCCGTCCCGGGGCATCATCCAGTGTATGGCGATGACAGCAGTACAGGGCGGCCA--      | 1825 |
| Db | 3672 | --CCGTCCTGGAAGTAAATCCGTTATATGTTGATGAGAACAGTGAAGTACTCCA         | 3729 |
| QY | 1826 | -GCAGTTTCACTCCCTACTGCTCCATGGCCCGCAGGCCAGCTATGCTTCCATGGGACCGC   | 1884 |
| Db | 3730 | GGGACATTTATACCTACTGTTTCAATGGCACATGCACAGCTTTGTTCCACGGGACCGG     | 3789 |
| QY | 1885 | GATGCGGTGAAGTTCTTGTCTCGTGTCCAGGGAACGTGTCTGGCCACCTGATGGCAGT     | 1944 |
| Db | 3790 | GATGCTGTGAAATTTCTTGTGGCAGTCCCAGGTGAGTCAATTAGCCCAAAAGTAGAGT     | 3849 |
| QY | 1945 | GTGCTGGACAGCCAGCGGAGGCGCTGGGCCAGCTGCCCTCGGAGGTGAGGGC           | 2004 |
| Db | 3850 | GGTGGGCGAGATCTTACAGCTGAC---AAGGCAGGTCATCTGCACAGGAGCCAGTAGC     | 3906 |
| QY | 2005 | CAGAAAGC---TGCGGAACGTGTGTGTGCTGAGCGGGGGGAGGGGTACATCGACTTCCGC   | 2061 |
| Db | 3907 | CAGACGCCCTTGAAGTCCATGTGTGTATCATGAGTGGAGAGGGGTACATTTGACTTCCGG   | 3966 |
| QY | 2062 | ATTGGAGACGGAGAGGACGACAGACGGAGGGGCGCAGGGGACATGAGCCAGGTGAAG      | 2121 |
| Db | 3967 | ATGGGTGATGAAGGTGGAGA-----ATCTGAACCTTCTGGGAGAGGATCTTCCACTGGAA   | 4020 |







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2004, 06:31:25 ; Search time 3786 Seconds  
(without alignments)  
4792.768 Million cell updates/sec

Title: US-10-019-495-26  
Perfect score: 3694  
Sequence: 1 ctggcaggttcttagtgagc.....gtcgtataaaaaaaaaa 3694

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3228839 seqs, 2456066551 residues

Total number of hits satisfying chosen parameters: 6457678

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB2.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID                     | Description       |
|------------|--------|-------------|--------|------------------------|-------------------|
| 1          | 3484.6 | 94.3        | 4701   | 9 US-09-864-864-326    | Sequence 326, App |
| 2          | 486.2  | 13.2        | 4688   | 13 US-10-276-774-804   | Sequence 804, App |
| 3          | 477    | 12.9        | 4286   | 14 US-10-044-090-549   | Sequence 549, App |
| 4          | 360.6  | 9.8         | 3791   | 15 US-10-240-965-3     | Sequence 3, Appli |
| 5          | 242.4  | 6.6         | 1413   | 9 US-09-925-297-167    | Sequence 167, App |
| 6          | 241.8  | 6.5         | 310    | 13 US-10-085-783A-1790 | Sequence 1790, Ap |
| 7          | 241.8  | 6.5         | 310    | 16 US-10-242-535A-1790 | Sequence 1790, Ap |
| 8          | 226    | 6.1         | 554    | 16 US-10-264-237-807   | Sequence 807, App |
| 9          | 219.4  | 5.9         | 675    | 9 US-09-917-800A-965   | Sequence 965, App |
| 10         | 189    | 5.1         | 189    | 9 US-09-796-692-6852   | Sequence 6852, Ap |
| 11         | 189    | 5.1         | 189    | 15 US-10-040-862-6852  | Sequence 6852, Ap |
| 12         | 189    | 5.1         | 189    | 16 US-10-057-475B-6852 | Sequence 6852, Ap |
| 13         | 189    | 5.1         | 189    | 16 US-10-154-884B-6852 | Sequence 6852, Ap |
| 14         | 116.6  | 3.2         | 6289   | 9 US-09-954-456-819    | Sequence 819, App |

|    |       |     |         |                        |                    |
|----|-------|-----|---------|------------------------|--------------------|
| 15 | 116.6 | 3.2 | 6289    | 9 US-09-954-456-1216   | Sequence 1216, Ap  |
| 16 | 116.6 | 3.2 | 6289    | 10 US-09-960-706-460   | Sequence 460, App  |
| 17 | 116.6 | 3.2 | 6289    | 10 US-09-873-319-280   | Sequence 280, App  |
| 18 | 116.6 | 3.2 | 7540    | 13 US-09-918-715-203   | Sequence 203, App  |
| 19 | 98.2  | 2.7 | 2160    | 14 US-10-044-090-704   | Sequence 704, App  |
| 20 | 90.4  | 2.4 | 505     | 16 US-10-264-237-316   | Sequence 316, App  |
| 21 | 90    | 2.4 | 2523    | 15 US-10-205-219-158   | Sequence 158, App  |
| 22 | 77.8  | 2.1 | 3476    | 16 US-10-104-047-841   | Sequence 841, App  |
| 23 | 77.8  | 2.1 | 3764    | 13 US-10-114-270-61    | Sequence 61, Appli |
| 24 | 77.8  | 2.1 | 5448    | 16 US-10-062-674-2053  | Sequence 2053, Ap  |
| 25 | 72.6  | 2.0 | 454     | 9 US-09-864-761-11580  | Sequence 11580, A  |
| 26 | 70    | 1.9 | 433     | 9 US-09-960-352-11926  | Sequence 11926, A  |
| 27 | 65    | 1.8 | 427     | 9 US-09-960-352-11694  | Sequence 11694, A  |
| 28 | 61    | 1.7 | 135638  | 15 US-10-314-657-1     | Sequence 1, Appli  |
| 29 | 59.8  | 1.6 | 1494    | 17 US-10-437-963-70080 | Sequence 70080, A  |
| 30 | 59.6  | 1.6 | 177     | 9 US-09-864-761-28166  | Sequence 28166, A  |
| 31 | 59.2  | 1.6 | 2787    | 15 US-10-156-761-6585  | Sequence 6585, Ap  |
| 32 | 59.2  | 1.6 | 9025608 | 15 US-10-156-761-1     | Sequence 1, Appli  |
| 33 | 57.6  | 1.6 | 65140   | 17 US-10-203-295-1     | Sequence 1, Appli  |
| 34 | 57.6  | 1.6 | 125401  | 17 US-10-203-295-35    | Sequence 35, Appl  |
| 35 | 55.2  | 1.5 | 1836    | 13 US-10-387-336-47    | Sequence 47, Appl  |
| 36 | 55.2  | 1.5 | 1944    | 13 US-10-387-336-46    | Sequence 46, Appl  |
| 37 | 55.2  | 1.5 | 2025    | 13 US-10-387-336-45    | Sequence 45, Appl  |
| 38 | 55.2  | 1.5 | 2352    | 9 US-09-476-242-26     | Sequence 26, Appl  |
| 39 | 55.2  | 1.5 | 2453    | 15 US-10-241-009-22    | Sequence 22, Appl  |
| 40 | 55.2  | 1.5 | 2453    | 15 US-10-190-434B-22   | Sequence 22, Appl  |
| 41 | 55.2  | 1.5 | 2652    | 17 US-10-441-926-1     | Sequence 1, Appli  |
| 42 | 55.2  | 1.5 | 2652    | 17 US-10-441-949-1     | Sequence 1, Appli  |
| 43 | 55    | 1.5 | 2298    | 9 US-09-476-242-22     | Sequence 22, Appl  |
| 44 | 55    | 1.5 | 2298    | 9 US-09-476-242-23     | Sequence 23, Appl  |
| 45 | 55    | 1.5 | 2298    | 9 US-09-476-242-24     | Sequence 24, Appl  |

ALIGNMENTS

RESULT 1  
US-09-864-864-326  
; Sequence 326, Application US/09864864  
; Patent No. US20020102679A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan L.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Fling, Steve P.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Carter, Darrick  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER  
; FILE REFERENCE: 210121.523  
; CURRENT APPLICATION NUMBER: US/09/864,864  
; CURRENT FILING DATE: 2001-05-23  
; NUMBER OF SEQ ID NOS: 341  
; SOFTWARE: Corixa Invention Disclosure Database  
; SEQ ID NO 326  
; LENGTH: 4701  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-864-864-326

Query Match 94.3%; Score 3484.6; DB 9; Length 4701;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 3490; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 180 GCCTCCCTGCTCCCTGCAGCTGAGTCCCAACGGGGCCAGGACACCGGATGAAGAA 239  
DB 1203 GCCGCCCAAGTACAGCAGCTGAGTCCCAACGGGGCCAGGACACCGGATGAAGAA 1262

|    |      |             |            |            |            |           |           |            |          |      |
|----|------|-------------|------------|------------|------------|-----------|-----------|------------|----------|------|
| QY | 240  | CGTGCCGGTGC | CGGTGTACTG | CCGCCCTCT  | GGTGGAGAA  | GGACCC    | CCCATGAAG | CTGTG      | 299      |      |
| Db | 1263 | CGTGCCGGTGC | CGGTGTACTG | CCGCCCTCT  | GGTGGAGAA  | GGACCC    | CCCATGAAG | CTGTG      | 1322     |      |
| QY | 300  | GTGTCCCGCGG | CGTCAACT   | GAGCGGTG   | GAGGCCCA   | ATGAGGAC  | GACGCTGGG | AAATGG     | 359      |      |
| Db | 1323 | GTGTCCCGCGG | CGTCAACT   | GAGCGGTG   | GAGGCCCA   | ATGAGGAC  | GACGCTGGG | AAATGG     | 1382     |      |
| QY | 360  | AGTCAAGCCAG | CGCCAGG    | CCCGGATCCC | CTGACCTG   | CGACCG    | GAAGGAG   | ACGGCGAGCC | 419      |      |
| Db | 1383 | AGTCAAGCCAG | CGCCAGG    | CCCGGATCCC | CTGACCTG   | CGACCG    | GAAGGAG   | ACGGCGAGCC | 1442     |      |
| QY | 420  | CAAGAGCGCC  | CACAGTCT   | CCCGAGAA   | GAGGCCCA   | AGGAGTCC  | CTGAAATG  | GACGC      | 479      |      |
| Db | 1443 | CAAGAGCGCC  | CACAGTCT   | CCCGAGAA   | GAGGCCCA   | AGGAGTCC  | CTGAAATG  | GACGC      | 1502     |      |
| QY | 480  | CACCTCCAGC  | CGGGTGTGG  | ATCCTGAC   | CAGACCTG   | ACCACAG   | CAAGTGGT  | GATCAT     | 539      |      |
| Db | 1503 | CACCTCCAGC  | CGGGTGTGG  | ATCCTGAC   | CAGACCTG   | ACCACAG   | CAAGTGGT  | GATCAT     | 1562     |      |
| QY | 540  | CGACGCCAAC  | CAGCGGCA   | CGGTGGTGA  | CCAGTTCAC  | CGTCTGCA  | ACGCGCAC  | GTGCT      | 599      |      |
| Db | 1563 | CGACGCCAAC  | CAGCGGCA   | CGGTGGTGA  | CCAGTTCAC  | CGTCTGCA  | ACGCGCAC  | GTGCT      | 1622     |      |
| QY | 600  | GTGCATCTCC  | AGCATCCCC  | CGGCGCAG   | CGACAGGACT | ACCCTCCC  | CGGGGAG   | ATGTTCT    | 659      |      |
| Db | 1623 | GTGCATCTCC  | AGCATCCCC  | CGGCGCAG   | CGACAGGACT | ACCCTCCC  | CGGGGAG   | ATGTTCT    | 1682     |      |
| QY | 660  | GGACAGCGAC  | GTGAACCC   | CAGAGGAC   | CCCGGCGC   | AGATGGCG  | TGCGCGGT  | ATCACCCCT  | 719      |      |
| Db | 1683 | GGACAGCGAC  | GTGAACCC   | CAGAGGAC   | CCCGGCGC   | AGATGGCG  | TGCGCGGT  | ATCACCCCT  | 1742     |      |
| QY | 720  | GGTGGGCTGT  | GCCACCCG   | CTGCAACGT  | CCCGGAGCA  | ACTGCTCCT | CCCGAGG   | GAGACAC    | 779      |      |
| Db | 1743 | GGTGGGCTGT  | GCCACCCG   | CTGCAACGT  | CCCGGAGCA  | ACTGCTCCT | CCCGAGG   | GAGACAC    | 1802     |      |
| QY | 780  | CCAGTGTCTA  | GACAGGGG   | CAGGGGAG   | GTGGCCACC  | ATCGCCRA  | CGGGAAG   | GTCAACCC   | 839      |      |
| Db | 1803 | CCAGTGTCTA  | GACAGGGG   | CAGGGGAG   | GTGGCCACC  | ATCGCCRA  | CGGGAAG   | GTCAACCC   | 1862     |      |
| QY | 840  | GTCCAGTCCA  | CAGAGGAG   | CCACAGAG   | GCCACG     | GAGTGC    | CAGACCT   | TGGCCACG   | 899      |      |
| Db | 1863 | GTCCAGTCCA  | CAGAGGAG   | CCACAGAG   | GCCACG     | GAGTGC    | CAGACCT   | TGGCCACG   | 1922     |      |
| QY | 900  | GCCAGAGACA  | GGCCACAT   | TGCGGCCG   | CGGCTCTC   | ACAGAGCA  | CGTTTCA   | CTGACCCAGC | 959      |      |
| Db | 1923 | GCCAGAGACA  | GGCCACAT   | TGCGGCCG   | CGGCTCTC   | ACAGAGCA  | CGTTTCA   | CTGACCCAGC | 1982     |      |
| QY | 960  | CCCGACCCG   | TCCTCTGG   | CCCCCAG    | CTGGCAG    | CGAGAAC   | GGGCCAG   | AGCCTGAC   | AGCAG    | 1019 |
| Db | 1983 | CCCGACCCG   | TCCTCTGG   | CCCCCAG    | CTGGCAG    | CGAGAAC   | GGGCCAG   | AGCCTGAC   | AGCAG    | 2042 |
| QY | 1020 | CAGCACGGC   | CCAGAGCC   | ACGCGGG    | ACCCCA     | CGGAGC    | AGCAGC    | AGTGTGC    | 1079     |      |
| Db | 2043 | CAGCACGGC   | CCAGAGCC   | ACGCGGG    | ACCCCA     | CGGAGC    | AGCAGC    | AGTGTGC    | 2102     |      |
| QY | 1080 | ACCCACCAT   | GTGGCTGG   | AGCCCC     | CAGACGG    | CTGCTAT   | TGTGCA    | CTCGGCTGT  | GSCCAA   | 1139 |
| Db | 2103 | ACCCACCAT   | GTGGCTGG   | AGCCCC     | CAGACGG    | CTGCTAT   | TGTGCA    | CTCGGCTGT  | GSCCAA   | 2162 |
| QY | 1140 | CTGGAAGA    | AGTGCC     | CTGCACTC   | CAAGTGA    | AGGATTCT  | TGTGCT    | GAGCCTGGT  | GCAATG   | 1199 |
| Db | 2163 | CTGGAAGA    | AGTGCC     | CTGCACTC   | CAAGTGA    | AGGATTCT  | TGTGCT    | GAGCCTGGT  | GCAATG   | 2222 |
| QY | 1200 | CAAAGGCG    | GTGTGGT    | CTGTGG     | CGAGCGG    | ACCTTCC   | ACCGTGG   | TGA        | 1259     |      |
| Db | 2223 | CAAAGGCG    | GTGTGGT    | CTGTGG     | CGAGCGG    | ACCTTCC   | ACCGTGG   | TGA        | 2282     |      |
| QY | 1260 | AGATGGCAG   | TGGATCTG   | AGCAACTA   | TACCTAA    | TGGACTG   | GGGCCAC   | CCGACCACTC | 1319     |      |
| Db | 2283 | AGATGGCAG   | TGGATCTG   | AGCAACTA   | TACCTAA    | TGGACTG   | GGGCCAC   | CCGACCACTC | 2342     |      |
| QY | 1320 | CATCCGCTG   | CAATGG     | CTGTTGT    | ACGACCG    | CGGTGTG   | GTGCTA    | CAAGAAC    | AAAGTGCA | 1379 |

|    |      |           |          |         |          |          |          |         |            |      |
|----|------|-----------|----------|---------|----------|----------|----------|---------|------------|------|
| Db | 2343 | CATCCGCTG | CAATGG   | CTGTTGT | TACGACCG | CGTGTG   | TGTGGCTA | CAAGAAC | AAAGTGCA   | 2402 |
| QY | 1380 | CGTCATCC  | AGCCCAAG | ACCATG  | CAGATAG  | AGAAGTCA | TTGACGCC | CAACCCG | CGCGGGA    | 1439 |
| Db | 2403 | CGTCATCC  | AGCCCAAG | ACCATG  | CAGATAG  | AGAAGTCA | TTGACGCC | CAACCCG | CGCGGGA    | 2462 |
| QY | 1440 | GAGCCAG   | TGCGG    | CAGCTG  | GCCTGG   | ATCGCG   | GATGGG   | TGTCCAT | CCGCCCTGGA | 1499 |
| Db | 2463 | GAGCCAG   | TGCGG    | CAGCTG  | GCCTGG   | ATCGCG   | GATGGG   | TGTCCAT | CCGCCCTGGA | 2522 |
| QY | 1500 | CTCCACC   | CTGAGG   | CTCTAC  | CATGCAC  | ACACG    | CACAGCA  | TCTAC   | AGGACG     | 1559 |
| Db | 2523 | CTCCACC   | CTGAGG   | CTCTAC  | CATGCAC  | ACACG    | CACAGCA  | TCTAC   | AGGACG     | 2582 |
| QY | 1560 | GCCCTAC   | GTCAAG   | ATGCTAG | GCACCTG  | GCAAGCT  | GGGTTCT  | CCTTC   | GTACG      | 1619 |
| Db | 2583 | GCCCTAC   | GTCAAG   | ATGCTAG | GCACCTG  | GCAAGCT  | GGGTTCT  | CCTTC   | GTACG      | 2642 |
| QY | 1620 | GGCCCTG   | CTTGC    | GCGG    | CAGCGG   | CTCTG    | GGTGG    | CACCGG  | CAACGG     | 1679 |
| Db | 2643 | GGCCCTG   | CTTGC    | GCGG    | CAGCGG   | CTCTG    | GGTGG    | CACCGG  | CAACGG     | 2702 |
| QY | 1680 | CATCCCC   | CTGAC    | GAGACTG | TGCTCT   | GCACCG   | AGGCCAG  | CTCTCG  | GGGCTCC    | 1739 |
| Db | 2703 | CATCCCC   | CTGAC    | GAGACTG | TGCTCT   | GCACCG   | AGGCCAG  | CTCTCG  | GGGCTCC    | 2762 |
| QY | 1740 | TAAGACAT  | CCCCAC   | CTCTG   | GGGAGG   | CGCCCG   | TCCCGG   | GGCAT   | CAACG      | 1799 |
| Db | 2763 | TAAGACAT  | CCCCAC   | CTCTG   | GGGAGG   | CGCCCG   | TCCCGG   | GGCAT   | CAACG      | 2822 |
| QY | 1800 | CGATGAC   | AGCAGT   | GACAGG  | CGCGC    | CAGCAG   | CTTCA    | TCCCCT  | ACTGCT     | 1859 |
| Db | 2823 | CGATGAC   | AGCAGT   | GACAGG  | CGCGC    | CAGCAG   | CTTCA    | TCCCCT  | ACTGCT     | 2882 |
| QY | 1860 | CCAGCTAT  | GCTTCC   | ATGGG   | CACCG    | CATGCC   | GTGAA    | GTTC    | TTGCT      | 1919 |
| Db | 2883 | CCAGCTAT  | GCTTCC   | ATGGG   | CACCG    | CATGCC   | GTGAA    | GTTC    | TTGCT      | 2942 |
| QY | 1920 | CGTGTGG   | CCACCC   | CTGA    | ATGSC    | AGTGTG   | TGGAC    | CCCAG   | CCGAGG     | 1979 |
| Db | 2943 | CGTGTGG   | CCACCC   | CTGA    | ATGSC    | AGTGTG   | TGGAC    | CCCAG   | CCGAGG     | 3002 |
| QY | 1980 | TGCCCTG   | CTCG     | GAGG    | TCGAGG   | GCAGAA   | GCTG     | CGGAA   | ACG        | 2039 |
| Db | 3003 | TGCCCTG   | CTCG     | GAGG    | TCGAGG   | GCAGAA   | GCTG     | CGGAA   | ACG        | 3062 |
| QY | 2040 | GGAGGG    | CTACAT   | CGACTT  | CCGCA    | TTGGAG   | ACGCG    | AGAG    | ACG        | 2099 |
| Db | 3063 | GGAGGG    | CTACAT   | CGACTT  | CCGCA    | TTGGAG   | ACGCG    | AGAG    | ACG        | 3122 |
| QY | 2100 | AGGGAC    | ATGAGC   | CCAGG   | TGAAG    | CCCGT    | GTGTCC   | AAAGG   | CAGAG      | 2159 |
| Db | 3123 | AGGGAC    | ATGAGC   | CCAGG   | TGAAG    | CCCGT    | GTGTCC   | AAAGG   | CAGAG      | 3182 |
| QY | 2160 | GTGGCAG   | GTGCTT   | ACACCC  | CCGAG    | TGAAG    | CTGTG    | CCCTG   | CCCGAC     | 2219 |
| Db | 3183 | GTGGCAG   | GTGCTT   | ACACCC  | CCGAG    | TGAAG    | CTGTG    | CCCTG   | CCCGAC     | 3242 |
| QY | 2220 | AGGACCC   | CCGAC    | CCCTG   | ACCCCG   | CGCGG    | CCCGG    | CCCGG   | CCCGG      | 2279 |
| Db | 3243 | AGGACCC   | CCGAC    | CCCTG   | ACCCCG   | CGCGG    | CCCGG    | CCCGG   | CCCGG      | 3302 |
| QY | 2280 | CCCTCTT   | CTAA     | CCCTC   | TCAAC    | CTGC     | AGCTT    | CAC     | CTG        | 2339 |
| Db | 3303 | CCCTCTT   | CTAA     | CCCTC   | TCAAC    | CTGC     | AGCTT    | CAC     | CTG        | 3362 |
| QY | 2340 | GAGTGGG   | GGATG    | CGGATC  | AGTGG    | GAGGAG   | GAGG     | GAGG    | GAGG       | 2399 |
| Db | 3363 | GAGTGGG   | GGATG    | CGGATC  | AGTGG    | GAGGAG   | GAGG     | GAGG    | GAGG       | 3422 |
| QY | 2400 | AGATGCT   | CTCGG    | GACAG   | TTTCC    | CGG      | GAGT     | CTC     | CTG        | 2459 |

Db 3423 AGATGCTCTCGGACAGTTTCCCGGCAGCTCCTGGCCAGCTTCCAGCCAGAGTCTCTCA 3482

QY 2460 AGTCCAGGGCACCTTGGGCCCCAGCGCAGGCAGAAATCCGAGGTGGTCTCTGGCTCTACCCCTG 2519

Db 3483 AGTCCAGGGCACCTTGGGCCCCAGCGCAGGCAGAAATCCGAGGTGGTCTCTGGCTCTACCCCTG 3542

QY 2520 GGCCTCCTACTCCCCAGCACCCCTGGAGGAGGCAGGGGCTCCCGCGCCGAGGCTGCCT 2579

Db 3543 GGCCTCCTACTCCCCAGCACCCCTGGAGGAGGCAGGGGCTCCCGCGCCGAGGCTGCCT 3602

QY 2580 GCCCTGGGCCCCACCTCTGCTCATGTGCTCATGGGGCCACCCCTGCCCTCCTGGGCCCTCACTC 2639

Db 3603 GGCCTGGGCCCCACCTCTGCTCATGTGCTCATGGGGCCACCCCTGCCCTCCTGGGCCCTCACTC 3662

QY 2640 TGCCTAGGGGAGCTGGGCCAGGCACTAGCCCTTTGGCCAGGAGGTGGGCCCTCAGGCTGCC 2699

Db 3663 TGCCTAGGGGAGCTGGGCCAGGCACTAGCCCTTTGGCCAGGAGGTGGGCCCTCAGGCTGCC 3722

QY 2700 CAGGTGCTGCACCCAGCCGCGCTTCTCTGGGGCTTCCCGTCTCAAGCCTCTATCCT 2759

Db 3723 CAGGTGCTGCACCCAGCCGCGCTTCTCTGGGGCTTCCCGTCTCAAGCCTCTATCCT 3782

QY 2760 GTCTGTCCCCACCCAGCTGTCCCTGCCAGGAGCTGGCATAAAGCACGAGGCCCGG 2819

Db 3783 GTCTGTCCCCACCCAGCTGTCCCTGCCAGGAGCTGGCATAAAGCACGAGGCCCGG 3842

QY 2820 CTCCCTGGGCGAGCTGCTTGAGAACAGAGACTGTACCCCATCTGCCCCATGAGGCGAGG 2879

Db 3843 CTCCCTGGGCGAGCTGCTTGAGAACAGAGACTGTACCCCATCTGCCCCATGAGGCGAGG 3902

QY 2880 CTCTTGCCAGCCCGCTTCTGACCCGCTGTCCCGCCAGGCTCTGCTGGGCGAGACTCAC 2939

Db 3903 CTCTTGCCAGCCCGCTTCTGACCCGCTGTCCCGCCAGGCTCTGCTGGGCGAGACTCAC 3962

QY 2940 CTTGGAGGAGTGGGCCCTTGAGTCTCTGCTCCCTCCAGAAAGCCCGCCAGGGTGGGATTCTC 2999

Db 3963 CTTGGAGGAGTGGGCCCTTGAGTCTCTGCTCCCTCCAGAAAGCCCGCCAGGGTGGGATTCTC 4022

QY 3000 AGGCTGCCAGGGCAGGCCCGCCAGGCTCAGGAAGAAAGGGAGGCCCTTGGCCCTCTCCGGGAT 3059

Db 4023 AGGCTGCCAGGGCAGGCCCGCCAGGCTCAGGAAGAAAGGGAGGCCCTTGGCCCTCTCCGGGAT 4082

QY 3060 CAGTCTTAGACACAGGCTCAGCCTCAGGTTGATGGGGATGATGTCTCCCGGGCCTG 3119

Db 4083 CAGTCTTAGACACAGGCTCAGCCTCAGGTTGATGGGGATGATGTCTCCCGGGCCTG 4142

QY 3120 CCTCTGCACGGGGCTCCACGGAGCCAGCTCCAGACACGCTACTAAGTGCCTAGGGTT 3179

Db 4143 CCTCTGCACGGGGCTCCACGGAGCCAGCTCCAGACACGCTACTAAGTGCCTAGGGTT 4202

QY 3180 GCCCGCTGTGGCCTGTCTCCAGGGAGCAACAGAGAGGCCCAAGAGAGGCCCGCTGGG 4262

Db 4203 GCCCGCTGTGGCCTGTCTCCAGGGAGCAACAGAGAGGCCCAAGAGAGGCCCGCTGGG 4262

QY 3240 CTGAGGATGGAGCCCGCCAGCCGACTCCAGCCCGCAGAGGGCAGAGCCCAACCTTGA 3299

Db 4263 CTGAGGATGGAGCCCGCCAGCCGACTCCAGCCCGCAGAGGGCAGAGCCCAACCTTGA 4322

QY 3300 CTGCTCTCCCTGCCAGTGGGCTCTCTGGCCCTATTCTACCTTCCAGGCCCACTGCAC 3359

Db 4323 CTGCTCTCCCTGCCAGTGGGCTCTCTGGCCCTATTCTACCTTCCAGGCCCACTGCAC 4382

QY 3360 TCCTGTCTGGGAGGCCCTTATGAGGGCAGCCCGCCAGCCCGCAGCCCAACCCAGAGA 3419

Db 4383 TCCTGTCTGGGAGGCCCTTATGAGGGCAGCCCGCCAGCCCGCAGCCCAACCCAGAGA 4442

QY 3420 AGCAGATCTTGGGGAGCTGCCCCACAAGCCCGCTGGCCACCGAGGGCTGCAGCCGCT 3479

Db 4443 AGCAGATCTTGGGGAGCTGCCCCACAAGCCCGCTGGCCACCGAGGGCTGCAGCCGCT 4502

QY 3480 GCGCTGCCGGCTTCTCCCCACCAACCCCTGCCACCTCCACTGTGATGTATGTCCGCTCCCTC 3539

Db 4503 GCGCTGCCGGCTTCTCCCCACCAACCCCTGCCACCTCCACTGTGATGTATGTCCGCTCCCTC 4562

QY 3540 GTCTGTTCCCCCAGGATCTCGAAGTGACTCCGGGCTGAGCAGTGGGGCGGCTGGGGAGG 3599

Db 4563 GTCTGTTCCCCCAGGATCTCGAAGTGACTCCGGGCTGAGCAGTGGGGCGGCTGGGGAGG 4622

QY 3600 GGTGACGATTCTCCTCAGGCTTTGGCCCTGCAAGCAAAACCCACATATCTCTGTATGT 3659

Db 4623 GGTGACGATTCTCCTCAGGCTTTGGCCCTGCAAGCAAAACCCACATATCTCTGTATGT 4682

QY 3660 AATAAATGTCTTAACGTCTG 3678

Db 4683 AATAAATGTCTTAACGTCTG 4701

RESULT 2

US-10-276-774-804  
; Sequence 804, Application US/10276774  
; Publication No. US20040053245A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; APPLICANT: Tang, Y, Tom et al  
; TITLE OF INVENTION: No. US20040053245A1e1 Nucleic Acids and Polypeptides  
; FILE REFERENCE: 21272-030  
; CURRENT APPLICATION NUMBER: US/10/276,774  
; CURRENT FILING DATE: 2002-11-18  
; PRIOR APPLICATION NUMBER: 09/560,875  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 09/496,914  
; PRIOR FILING DATE: 2000-02-03  
; NUMBER OF SEQ ID NOS: 2700  
; SOFTWARE: Custom  
; SEQ ID NO 804  
; LENGTH: 4688  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-276-774-804

Query Match 13.2%; Score 486.2; DB 13; Length 4688;  
Best Local Similarity 58.5%; Pred. No. 2.7e-112;  
Matches 1191; Conservative 0; Mismatches 753; Indels 93; Gaps 16;

QY 214 GGCACGAGGACACGCGATGAAGAACCTGCGCGGTGCTGCGGTGCTGCGCGCTCTGTGTG 273

Db 1973 GGTCAAGGTGAAATTAAGATGAAATTTACCTGTGCTGTCTATCTCAGACCTCTGGAT 2032

QY 274 GAGAAGGACCCACCATGAAGCTGTGTTGTCGCGCGGTCAACCTGAGCGGTGGAGG 333

Db 2033 GAAAAGATACATCAATGAAGCTGTGTTGCTGCTGTTGAGTCAATTTATCTGTTGTTG 2092

QY 334 CCCAATGAGGACGACGCTGGGAATGGAG-----TCAAGCCAGCCAGGCCGC 381

Db 2093 ACCAGATGTTGTTCTGTTGTTGGAGCAAGTGTATTTTACAAGGATGTTGCTGGTTG 2152

QY 382 GATCCCTGACCTGCGACCGCGAAGGAGACGGCGAGCCCAAGAGCGCCCAACACGTTCTCC 441

Db 2153 GATACAGAAGGCAGTAAACAGCGAAGTGCCTCTCAGAGTAGTTTAGATAAGTTAGATCAG 2212

QY 442 GAG---AAGAAGAAGCCCAAGGAGTCCCTGAAATGAGACGCCACCTCCAGCCGGTGTGG 498

Db 2213 GAACTTAAGGAACAGCAGAGGAGTTAAAAAATCAAGAAGATTATCCAGTCTAGTTGG 2272

QY 499 ATCTGACCCAGCACCTGACCCAGCAGCAAGGTGGTGTGATCATCGACGCCAACCCAGCCGGC 558

Db 2273 ATCTGACCCAGCACCTGACCTCGGCTACAAAGTTCTATTATTGATGCTGTTCAACCTGGC 2332

QY 559 ACGTGGTGGACCACTTACCCGTCTGCAACCGCAGCTGCTGTGATCTCCAGATCCCC 618

Db 2333 AACATCCTAGACAGTTTCACTGTTGCAACTCTCATGTTCTGTCATTTGATGCTGTTCAACCTGGC 2392

QY 619 GCGGCCAGGACAGGACTACCTCCCTCCCGG--GGAGATGTTCTCTGACAGCGACGTGAACCC 677

Db 2393 GGTGACAGGAAACAGACTACCTCTGACGAGAGAGATCTTTTCAAGATCTGTTGTCAGTAGAC 2452

QY 678 AGAGGACC-----CGGGCGCAGATGGCGTGCTG 705  
Db 2453 AAAGCATCTTTATGTGGAAGTATGACAAGCAACAGCTCAGCAGAGACAGACAGCCTGTGA 2512  
QY 706 GCCGGTATCACCCCTGGTGGCTGTGC--CACCCGCTGCAACGTGCCGCGGAGCAACTGCT 763  
Db 2513 GGAGGCATCAGATGGTTGGTTGTCTGCGAAGGTGTACGGGAGCTGCCACTTCCCT 2572  
QY 764 CCTCCGAGGGGACACC-CCAGTGTCTAGACAAGGGGCGAGGGGAGGTGGCCACCATCGCC 822  
Db 2573 AGTACAAATCGTGCTTCTCCAGTGTGATTAACCAACAGAAATGGAAGCAGAAAATAGT 2632  
QY 823 AACGGGAAGGTCAACCCGTCAGTCCAGTCCACAGAGGAGGCCACAGAGGTCGCCA 882  
Db 2633 GA--GGTTGATGAAAATGTTCCAACAGCAGAGAAAAGCAACTGAAGCTACAGAAGGG--A 2688  
QY 883 GACCCTGGGCCAGCGAGCCAGACAGACAGCCACATTTGGCGCCCGGGCCTCTCACAGAGCAC 942  
Db 2689 ATGCGGGGTGAGTGAAGACACAGTTGGACATCTCCCAAACTGGCGTCTACACAGAGCAT 2748  
QY 943 GTCTTCACTGACCCAG-----CCCCGACCCCTCTCTGGCCCCCAGCC 986  
Db 2749 GTCTTTACAGATCCTTTGGGAAGTTGAGATCCCAAGAACCTCTCCCCAGTGTATCAGTC 2808  
QY 987 TGGCAGCGAGAAACGGGCCAGAGCCTGACAGCAGCAGCACACGGCCA---GAGCCAGAGCC 1043  
Db 2809 GAGCAATGACTCAGATGCATATAAAGATCAATATCAGTACTGCCAAATGAACAAGACTT 2868  
QY 1044 CAGCGGGACCCCGGAGCAGCAGCAGTGTGACACCCACCATGTGGCTGGAGCCCCA 1103  
Db 2869 GGTGAGAGAAGAAGCCAGAAAATGAGTAGTCTTTTACCAACTATGTGGCTTGGAGCTCA 2928  
QY 1104 GAACGGCTGGCTCTATGTGCACTCGGCTGTGGCCAACTGGAAGAGTGCCTGCACCTCCAT 1163  
Db 2929 AAATGGCTGTTTGTATGTCCATTCTGTAGCCAGTGGAGGAAATGTCTCCATTCCAT 2988  
QY 1164 CAAGCTGAAGGATCTGTGCTGAGCCTGGTGTCATGTCACAAAGGCCGTGTGCTGGTGTCT 1223  
Db 2989 TAAACTTAAAGATTCGATTCTCAGTATTGTACACGTGAAGGGAATCGTGTAGTAGCCCT 3048  
QY 1224 GGCGACGGGACCCCTGGCCATCTTCCACCGTGGTGAAGATGGCCAGTGGGATCTGACAA 1283  
Db 3049 GGCTACGGCACCCCTTGCAATCTTTCACAGAGGAGTGGATGGCAGTGGGATTTGTCAAA 3108  
QY 1284 CTATCACCTAATGGACCTGGGCCACCCGACCACTCCATCCGCTGCATGGCTGTTGTGTA 1343  
Db 3109 CTATCACCTCTTAGACCTTGGACCGCTCATCATTCATCCGTTGCATGACTGTGGTACA 3168  
QY 1344 CGACCGCGTGTGGTGTGGCTACAAGAACAAAGGTGCACGTTCATCCAGCCCAAGACCATGCA 1403  
Db 3169 TGACAAAGTCTGGTGTGGCTATAGGAACAAAATCTATGTGGTGACGCCAAAGGCCATGAA 3228  
QY 1404 GATAGAGAAGTCATTTGACGCCCCACCCGCGGGGAGAGCCAGGTGCGCAGCTGGCGTG 1463  
Db 3229 AATAGAGAAAATCTTTTATGTCACATCCAGGAAGGAGAGCCAAAGTGCACAGCTTGCCTG 3288  
QY 1464 GATCGGCGATGGCGTATGGGTGTCCATCCGCTGGACTCCACCTTGAGGCTCTACCATGC 1523  
Db 3289 GGTGGGGATGGCGTGTGGGTCTCCATCGCTTGGATTCTACGCTCCGTTCTATCATGC 3348  
QY 1524 ACACCGCACAGCATCTACAGGACGTGGACATTGAGCCCTACGTGAGCAAGATGCTAGG 1583  
Db 3349 ACACACTTATCAACATCTACAGGATGTGGACATTGAGCCCTATGTAAGCAAAATGTTAGG 3408  
QY 1584 CACTGGCAAGCTGGGTTTCTCCTTCTGTTACGCATCACGGCCCTGCTTGTGCGGGCAGCCG 1643  
Db 3409 TACTGGAAAACCTGGGCTTCTCTTTGTGAGAATTACAGCTTTATGGTGTCTGTAATCG 3468  
QY 1644 GCTCTGGGTGGGCACCGCAACGGAGTGGTGTCTCCATCCCCCTGACAGAGACTGTGTT 1703  
Db 3469 TTTGTGGGTGGGACAGGAAATGGTGTCTATCTCCATCCCATTTGACAGAAACCGTAAT 3528  
QY 1704 CTGCAACGAGCCAGCTCTCTGGGCTCCGAGCCAAATAAGACATCCCCCACCTCTGGGGA 1763

Db 3529 CCTCCACCAGGGACGTTTACTGGGGCTGAGGGCAAAATAAAACCTCAGGTGTACCAGGAAA 3588  
QY 1764 GGGCGCCCGTCCCGGGGCATCATCCACGTGTATGGCGATGACAGCAGTGACA---GGGC 1820  
Db 3589 T-----CGTCTGGAAGTGTAAATCCGTGTATATGGTATGATAAACAAGTAAAGTGAC 3642  
QY 1821 GGCCAGCAGCTTCATCCCTACTGTCTCCATGGCCCCAGGCCAGCTATGCTTCCATGGCA 1880  
Db 3643 TCCAGGGACATTTATACCCCTATTTCAATGGCACATGCACAGCTTTGCTTCCATGGCA 3702  
QY 1881 CCGCGATGCCGTGAAGTCTTTGTCTCGGTGCCAGGGAACGTGCTGGCCACCCCTGAATGG 1940  
Db 3703 CCGGATGCTGTGAAATCTTTGTGGCAGTCCAGGTCAAGTCAATCAGCCCCAAGTAG 3762  
QY 1941 CAGTGTCTGGACAGCCCCAGCGAGGGCCCTGGGCCAGCTGCCCTGCCTCGAGGTCGA 2000  
Db 3763 CAGT-AGTGGCACGGATCTGACGGGTGACAAAGGCAGGGGCCATCTGCACAGGAGCCTGG 3821  
QY 2001 GGGCCAGAAAGC---TGCGAAGTGTGTGGTGTGAGCGGGCGGAGGGCTACATCGACTT 2057  
Db 3822 TAGTCAGACGCGCCTTGAAGTCTATGCTTGTCTCAGTGGAGGAGGGCTACATCGACTT 3881  
QY 2058 CCGCATTTGGACGAGAGGAGGACGACGAGACGGAGGGCGGCGAGGGGACATGAGCCAGGT 2117  
Db 3882 CCGAATGGTGTATGAAG-----GTGGAGAAATCAGAACTTCTTGGAGAGGATCTTCCACTT 3936  
QY 2118 GAAGCCCGTGTGTCTCAAGGCGAGCGCAGTCAATCATCATCTGTTGGCAGGTGTCTTA 2174  
Db 3937 GAACCTTCTG-TCACCAAAGCAAGAAAGGAGTCACTTGATAGTGTGGCAAGTGATGTA 3992

RESULT 3  
US-10-044-090-549  
; Sequence 549, Application US/10044090  
; Publication No. US20020137081A1  
; GENERAL INFORMATION:  
; APPLICANT: Olga Bandman  
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION  
; FILE REFERENCE: PA-0028 US  
; CURRENT APPLICATION NUMBER: US/10/044,090  
; CURRENT FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 850  
; SOFTWARE: PERL Program  
; SEQ ID NO 549  
; LENGTH: 4286  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1327121.3  
US-10-044-090-549

Query Match 12.9%; Score 477; DB 14; Length 4286;  
Best Local Similarity 56.5%; Pred. No. 5.5e-110;  
Matches 1154; Conservative 0; Mismatches 805; Indels 84; Gaps 11;

QY 214 GGCCAGGAGGACACGCGGATGAAGAACGTGCCGTGCCGCGGTGACTGCCGCCCTCTGGTG 273  
Db 1030 GGTCAAGGTGAAATAAGATGAAAAATTTACCTGTGCCTGTCTATCTCAGACCTCTGGAT 1089  
QY 274 GAGAAAGGACCCACCACCATGAAGCTGTGGTGTGCCGCGGGCGTCAACCTGAGCGGGTGGAGG 333  
Db 1090 GAAAAAGATACATCAATGAAGCTGTGGTGTGCTGTTGGAGTCAATTTATCTGGTGGGAAG 1149  
QY 334 CCCAATGAGGACGACGCTGGGAATGGAG-----TCAAGCCAGCGCCAGGCCGC 381  
Db 1150 ACCAGAGATGGTGGTTCTGTTGTTGGAGCAAGTGTATTTTACAAGGATGTTGCTGTTTG 1209  
QY 382 GATCCCTGACCTGCGACCGCAAGGAGAGCGGCGAGCCCAAGAGCGCCACACGTCTCCC 441  
Db 1210 GATACAGAGGCAGTAACACAGCGAAGTGCCTCTCAGAGTAGTTTAGATAAGTTAGATCAG 1269

|    |      |  |      |
|----|------|--|------|
| QY | 442  | GAG--AAGAAAGAGGCAAGGAGCTCCCTGAAATGGACGCCACCTCCAGCGGGTGTGG      | 498  |
| Db | 1270 | GAACCTTAAGGAACAGCAGAAAGGATTAAAAAATCAAGAAGAATTATCCAGTCTAGTTGG   | 1329 |
| QY | 499  | ATCCTGACCAGCACCTTGACACCAGCAAGGTGGTGTATCATCGACGCCAACCAAGCCGGC   | 558  |
| Db | 1330 | ATCTGTACCAGCACTCATTCGGCTACAAAAGTTCTTATTTATGATGCTGTTCACCTGGC    | 1389 |
| QY | 559  | ACGGTGGTGGACCAAGTTCAACCGTCTGCAACGGCAGCTGCTGTGTCATCTCCAGCATCCCC | 618  |
| Db | 1390 | AACATCCTAGACAGTTCACTGTTTGCAACTCTCATGTTCTGTGTCATTGCAAGTGTGCCA   | 1449 |
| QY | 619  | GCGGCCAGCGACAGCGACTACCCCTCCCGG-GGAGATGTTCTCTGGACAGCGACCTGAACCC | 677  |
| Db | 1450 | GGTGACGAGAAAAGACTACCCCTGCAGGAGAAAGATCTTTCAGAAATCTGGTCAGGTAGAC  | 1509 |
| QY | 678  | AGAGGACC-----CGGGCGCAGATGGCGTGGCTG                             | 705  |
| Db | 1510 | AAAGCATCTTTATGTGGAAGTATGACAAGCAACAGCTCAGCAGAGACAGACAGCCTGTTA   | 1569 |
| QY | 706  | GCCGGTATCACCTCGTGGCTGTGCCACCCGCTGCAACGTGCCGGAGCAACTGCTCC       | 765  |
| Db | 1570 | GGAGGCATCACAGTGGTTGGTTGTTCTGCAGAAAGTGTGACGGGAGCTGCCACTTCCCT    | 1629 |
| QY | 766  | TCCGAGGGGACACCCAGTGTGTAGACAAGGGCAGGGGAGGTGGCCACCATCGCCAAAC     | 825  |
| Db | 1630 | AGTACAAATGGTGTCTTCCAGTGATGGATAAACACACAGAAATGGAAGCAGAAAATAGT    | 1689 |
| QY | 826  | GGGAAGTCAACCCGTCCAGTCCACAGAGGAGGCCACAGAGGCCACGGAGGTGCCAGAC     | 885  |
| Db | 1690 | GAGTTGATGAAAAATGTTCCAAACAGCAGAGAAAGCAAACTGAAGCTACAGAAGGGA--AT  | 1746 |
| QY | 886  | CCTGGGCCACGAGCCAGAGACAGCCACATTGGGCCCGGGCCTCTCACAGAGCAGCTC      | 945  |
| Db | 1747 | GCGGGTCACTGAAGACACAGTGGACATCTCCCAAACCTGGCGTCTACACAGAGCATGTC    | 1806 |
| QY | 946  | TTCACTGACCC-----AGCCCCGACCCCGTCTCTGGCCCCCAGCCTGGC              | 990  |
| Db | 1807 | TTTACAGATCCTTTGGGAGTTTCAGATCCCAGAAACCTCTCCCGAGTGTATCAGTCGAGC   | 1866 |
| QY | 991  | AGCGAGAACGGCCAGAGCCTTGACAGCAGCAGCAGCAGCCCA---GAGCCAGAGCCAGC    | 1047 |
| Db | 1867 | AATGACTCAGATGCATATAAAGATCAAATATCAGTACTGCCAAATGAACAAGACTTGGTG   | 1926 |
| QY | 1048 | GGGACCCACGGGAGCAGGCAGCAGTGTGTGACCCCAACATGTGGCTGGAGCCAGAAAC     | 1107 |
| Db | 1927 | AGAGAAGACCCAGAAAATGAGTAGTCTTTTACCAACTATGTGGCTTGGAGCTCAAAAT     | 1986 |
| QY | 1108 | GGCTGGCTCTATGTGCACTCGGCTGTGGCCAACTGGAAAGATGCCCTGCACTCCATCAAG   | 1167 |
| Db | 1987 | GGCTGTTGTATGTCCATTCTGTAGCCAGTGGAGGAAATGTCTCCATTCCATTAAA        | 2046 |
| QY | 1168 | CTGAAGGATCTGTGTGAGCCTGGTGCATGTCAAAGCCGTGTGCTGGTGGCTCTGGCG      | 1227 |
| Db | 2047 | CTTAAAGATTTCGATTCTCAGTATTGTACACGTGAAGGGAATCGTGTATTAGCCCTGGCT   | 2106 |
| QY | 1228 | GACGGACCCCTGGCCATCTTCCACCGTGGTGAAGATGGCCAGTGGGATCTGAGCAACTAT   | 1287 |
| Db | 2107 | GACGACCCCTTGCAATCTTTTCACAGAGGAGTGGATGGGCAGTGGGATTTGTCAAACATAT  | 2166 |
| QY | 1288 | CACCTAATGGACCTGGGCCACCCGACCACTCCATCCGCTGCATGGCTGTGTGTGTACGAC   | 1347 |
| Db | 2167 | CACCTCTTAGACCTTGGACGGCCTCATCATTCATCCGTTGCATGACTGTGGTACATGAC    | 2226 |
| QY | 1348 | CGCGTGGTGTGGCTTACAAGAACAAAGGTGCACGTATCCAGCCCCAAGACCATGCAGATA   | 1407 |
| Db | 2227 | AAAGTCTGGTGTGGCTATAGGAACAAAATCTATGTGGTGCAGCCAAAGGCCATGAAAAATA  | 2286 |
| QY | 1408 | GAGAAGTCATTTTGACGCCCCACCCGCGGGAGAGCCAGGTGCGGACGTGGCGTGGATC     | 1467 |
| Db | 2287 | GAGAAATCTTTTGATGCACATCCCAGGAAGGAGGCCAAGTGCAGACAGTTGCGTGGGTG    | 2346 |
| QY | 1468 | GGCGATGGCGGTATGGTGTCCATCCGCCTGGACTCCACCCCTGAGGCTTACCATGACAC    | 1527 |

RESULT 4  
US-10-240-965-3  
; Sequence 3, Application US/10240965  
; Publication No. US20030165924A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.  
; APPLICANT: SHIFFMAN, Dov  
; APPLICANT: SOMOGYI, Roland  
; APPLICANT: LAWN, Richard M.  
; APPLICANT: SEILHAMER, Jeffrey J.  
; APPLICANT: PORTER, Gordon J.  
; APPLICANT: MIKITA, Thomas  
; APPLICANT: TAI, Julie  
; TITLE OF INVENTION: GENES EXPRESSED  
; FILE REFERENCE: PA-0025 PCT  
; CURRENT APPLICATION NUMBER: US/10/74  
; CURRENT FILING DATE: 2002-10-04  
; PRIOR APPLICATION NUMBER: 60/195,100  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 276  
; SOFTWARE: PERL Program

|   |  |      |  |      |
|---|--|------|--|------|
| ; SEQ ID NO 3   |  | 1048 | GGGACCCACGGGAGCAGGCAGCAGTGTGTCACCCACCATGTGGCTGGAGCCCAAG        | 1107 |
| ; LENGTH: 3791  |  |      |  |      |
| ; TYPE: DNA   |  | 2944 | AGAGAAGAACCCAGAAATGAGTAGTCTTTTACCAACTATGTGGCTTGAGCTCAAAAT      | 3003 |
| ; ORGANISM: Homo sapiens  |  |      |  |      |
| ; FEATURE:  |  | 1108 | GGCTGGCTCTATGTGCACCTGGCTGTGGCAACTGGAAGAAGTGCCTGCATCCATCAAG     | 1167 |
| ; NAME/KEY: misc feature  |  |      |  |      |
| ; OTHER INFORMATION: Incyte ID No. US20030165924A1 247178.2       |  | 3004 | GGCTGTTTGTATGTCCATTCATCTGTAGCCAGTGGAGAAATGTCTCCATCCATTA        | 3063 |
| US-10-240-965-3   |  |      |  |      |
| Query Match   |  | 1168 | CTGAAGGATTTCTGTGCTGAGCCTGGTGCATGTCAAAGGCCGTGTCTGGTGGCTCTGGCG   | 1227 |
| Best Local Similarity 9.8%; Score 360.6; DB 15; Length 3791;      |  |      |  |      |
| Matches 991; Conservative 0; Mismatches 669; Indels 119; Gaps 11; |  | 3064 | CTTAAAGATTCGATTCTCAGTATTGTACACGTGAAGGGAATCGTGTAGTACCCCTGGCT    | 3123 |
|   |  |      |  |      |
|   |  | 1228 | GACGGACCCCTGGCCATCTTCCACCGTGGTGAAGATGGCCACTGGGATCTGAGCAACTAT   | 1287 |
|   |  |      |  |      |
|   |  | 3124 | GACGGACCCCTGGCAATCTTTCACAGAGAGTGGATGGGCAATTTGTCAAACTAT         | 3183 |
|   |  |      |  |      |
|   |  | 1288 | CACCTAATGGACCTGGGCCACCCGCCACCACTCCATCCGCTGCATGGCTGTTGTGACGAC   | 1347 |
|   |  |      |  |      |
|   |  | 3184 | CACCTCTTAGACCTTGGACGGCCTCATCAATCCATCCGTTGCATGACTGTGGTACATGAC   | 3243 |
|   |  |      |  |      |
|   |  | 1348 | CGCGTGTGTGTGGCTACAAGAACAAAGGTGCACGTCTATCCAGCCCAAGACCATGAGATA   | 1407 |
|   |  |      |  |      |
|   |  | 3244 | AAAGTCTGGTGTGGCTATAGAAACAAATCTATGTGGTGCAGCAAGGCCATGAAATA       | 3303 |
|   |  |      |  |      |
|   |  | 1408 | GAGAAATCAATTGACGCCACCCGCCGGGAGAGCAGGTGCGGAGCAGTGGCTGGATC       | 1467 |
|   |  |      |  |      |
|   |  | 3304 | GAGAAATCTTTTGATGCACATCCAGGAAGGAGAGCAGCAAGTGGACA-CTTTCTGGGTG    | 3362 |
|   |  |      |  |      |
|   |  | 1468 | GCGCATGGCGTATGGGTGT-CCATCCGCTGGACTCCACCCCTGAGGCTCTACCATGCACA   | 1526 |
|   |  |      |  |      |
|   |  | 3363 | GGGATGGCGTGTGGGTCTGCCATTCGCTGGATTTCTACGCTCCGCTCTCTATCATGCACA   | 3422 |
|   |  |      |  |      |
|   |  | 1527 | CACGACCCAGCATCTACAGGACGTGGACATTTGAGCCCTACGTCAGCAAGATGCTAGGCAC  | 1586 |
|   |  |      |  |      |
|   |  | 3423 | CATTATCAACATACAGGATGTGGACATTTGAGCCTTATGTAAGCAAAATGTTAGGTAC     | 3482 |
|   |  |      |  |      |
|   |  | 1587 | TGGCAAGCTGGGTTCTCTTCTTGTGAGAAATTACAGCTCTTATGGTGTCTTGTAAATCGTTT | 1646 |
|   |  |      |  |      |
|   |  | 3483 | TGGAAACTGGGCTTCTCTTTTGTGAGAAATTACAGCTCTTATGGTGTCTTGTAAATCGTTT  | 3542 |
|   |  |      |  |      |
|   |  | 1647 | CTGGTGGGCACCGGCAACGGAGTGGTCACTCTCCATCCCTGACAGAGACTGTGTCCT      | 1706 |
|   |  |      |  |      |
|   |  | 3543 | GTGGTGGGCACAGGAAATGGTGTCAATATCTCCATCCCTGACAGAAAC-----          | 3593 |
|   |  |      |  |      |
|   |  | 1707 | GCACGAGGCCAGCTCTGGGCTCCGAGCCAAATAGACATCCCCACCTCTGGGAGGG        | 1766 |
|   |  |      |  |      |
|   |  | 3594 | -----AAATAAACCTCAGGTGTACCAGG                                   | 3617 |
|   |  |      |  |      |
|   |  | 1767 | CGCCCTCCGGGGCATCATCCAGTGTATGGCGATGACAGCAGTACACA---GGGCGGC      | 1823 |
|   |  |      |  |      |
|   |  | 3618 | AAATCGTCTGGAGTGTAAACCGTGTATATGGTATGAAACACAGTATAAAGTACTCC       | 3677 |
|   |  |      |  |      |
|   |  | 1824 | CAGCAGTTTCATCCCTACTGCTCCATGGCCCGAGCCAGCTATGCTTCCATGGCACCG      | 1883 |
|   |  |      |  |      |
|   |  | 3678 | AGGGACATTTATACCTTATTTCAATGGCACATGCACAGCTTTGCTTCCATGGCACCG      | 3737 |
|   |  |      |  |      |
|   |  | 1884 | CGATGCGGTGAAGTTCTTTGTCTCGGTGCCAGGGAACGT                        | 1922 |
|   |  |      |  |      |
|   |  | 3738 | GGATGCTGTGAAATTTCTTTGTGGCAGTCCAGGTCAAAGT                       | 3776 |
|   |  |      |  |      |
| RESULT 5  |  |      |  |      |
| US-09-925-297-167   |  |      |  |      |
| ; Sequence 167, Application US/09925297                           |  |      |  |      |
| ; Patent No. US20020081659A1                                      |  |      |  |      |
| ; GENERAL INFORMATION:  |  |      |  |      |
| ; APPLICANT: Rosen et al.   |  |      |  |      |
| ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies      |  |      |  |      |
| ; FILE REFERENCE: PA105   |  |      |  |      |
| ; CURRENT APPLICATION NUMBER: US/09/925,297                       |  |      |  |      |
| ; CURRENT FILING DATE: 2001-08-10                                 |  |      |  |      |
| ; PRIOR APPLICATION NUMBER: PCT/US00/05989                        |  |      |  |      |
| ; PRIOR FILING DATE: 2000-03-08                                   |  |      |  |      |

|    |      |   |      |
|----|------|---|------|
| QY | 214  | GGCCAGGAGGACACGCGGATGAAGAACGTGCCGCTGCCGCTGTACTGCCGCTCTCTGGTG  | 273  |
| Db | 2047 | GGTCAAGGTGAAATAAGATGAAATAATTTACCTGTGCTGTCTATCTCAGACCTCTGGAT   | 2106 |
|    |      |   |      |
| QY | 274  | GAGAAGGACCCACCATGAAGCTGTGGTGTGCCCGGGCGTCAACCTGAGCGGGTGGAG     | 333  |
| Db | 2107 | GAAAAAGATACATCAATGAAGCTGTGGTGTGGTGTGGAGTCAATTTATCTGGTGGGAG    | 2166 |
|    |      |   |      |
| QY | 334  | CCCAATGAGGACGACGCTGGGAATGGAG-----TCAAGCCAGCGCCAGGCCGC         | 381  |
| Db | 2167 | ACCAGAGATGGTGTCTGTGTTGGAGCAAGTGTATTTTACAGGATGTTGCTGGTTG       | 2226 |
|    |      |   |      |
| QY | 382  | GATCCCTGTACCTTGGACCGCGAAGGAGACGGCGAGCCCCAAGAGCGCCACACGTCTCCC  | 441  |
| Db | 2227 | GATACAGAAAGGCATPAAACAGCGAAGTGCCTCTCAGAGTAGTTTAGATAAGTTAGATCAG | 2286 |
|    |      |   |      |
| QY | 442  | GAG--AAGAAGAGGCCCAAGGAGCTCCCTGAAATGGAGCGCCACCTCCAGCGGGTGTGG   | 498  |
| Db | 2287 | GAACTTAAGGAACAGCAGAGGAGGTAAATAATCAAGAAAGAAATATCCAGTCTAGTTGG   | 2346 |
|    |      |   |      |
| QY | 499  | ATCCTGACCGACCTGTACCAACAGCAAGGTGGTGTATCATCGACGCCCAACAGCGGGC    | 558  |
| Db | 2347 | ATCTGTACCGACCTCATTCGGCTACAAAAGTTCTTATTTAGATGCTGTTCAACTGGC     | 2406 |
|    |      |   |      |
| QY | 559  | ACGGTGGTGACCGTTCACCGTGTGCAACGCGCAGTGTGTGATCTCCAGCATCCCC       | 618  |
| Db | 2407 | AACATCTAGACAGTTTCACTGTTTGCAACTCTCATGTTCTGTGCAATTGCAAGTGCCA    | 2466 |
|    |      |   |      |
| QY | 619  | GCGGCAGCGACGCGACTACCTCCCGG-GGAGATGTTCTTGACAGCGACGTGAACCC      | 677  |
| Db | 2467 | GGTGCAGAGAAACAGACTACCTTGAGGAGAGATCTTTCAGAAATCGGTACAGGTAGAC    | 2526 |
|    |      |   |      |
| QY | 678  | AGAGGACC-----CGGGCGCAGATGGCGTGTG                              | 705  |
| Db | 2527 | AAAGCATCTTTATGTGGAAGTATGACAAAGCAACAGTCTCAGCAGACAGACAGCCTGTTA  | 2586 |
|    |      |   |      |
| QY | 706  | GCCGTTATCACCTGTGGGCTGTGCCACCCGCTGCAACAGTGTGCCGCGAGCAACTGCTCC  | 765  |
| Db | 2587 | GGAGGATACAGTGTGGTGTGTTCTGCAAGAGTGTGACGGGAGCTGCCACTTCCCT       | 2646 |
|    |      |   |      |
| QY | 766  | TCCCGAGGGGACACCCCAAGTGTGACAAAGGGGAGGGGAGGTGGCCACCATCGCCAAC    | 825  |
| Db | 2647 | AGTACAAATGGTGTCTTCCAGTGTATGGATAAAACCAACAGAAATGGAAGCAGAAATAGT  | 2706 |
|    |      |   |      |
| QY | 826  | GGGAAGTCAACCCGTTCCAGTCCACAGAGGAGGCCACAGAGGCCACGGAGGTGCCAGAC   | 885  |
| Db | 2707 | GAGGTTGATGAAATGTTCCAAACAGCAGAGAAGCAACTGAAGCTACAGAAGG---AAT    | 2763 |
|    |      |   |      |
| QY | 886  | CCTGGGCCAGCGAGCCAGACAGCCACATTCGGCCCGGCTCTCTCAGAGCAACGTC       | 945  |
| Db | 2764 | GCGGGTCAAGTGAAGACACAGTGGACATCTCCCAACTGGCGTCTACACAGAGCATGTC    | 2823 |
|    |      |   |      |
| QY | 946  | TTCACTGACCC-----AGCCCGACCCCTGCTCTGCCCCCAGCCTGGC               | 990  |
| Db | 2824 | TTTACAGATCTTTGGGAGTTTCAGATCCCAAGAGACCTCTCCCAAGTGTATCAGTCGAGC  | 2883 |
|    |      |   |      |
| QY | 991  | AGCGAGAACGGGCGCAGAGCCTGACAGCAGCAGCACACGGCCA---GAGCCAGACCCAGC  | 1047 |
| Db | 2884 | AATGACTCAGATGCATATAAAGATCAAATATCAGTACTGCCAAATGAACAAGACTTGGTG  | 2943 |
|    |      |   |      |

; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 167  
; LENGTH: 1413  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1376)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-297-167

Query Match 6.6%; Score 242.4; DB 9; Length 1413;  
Best Local Similarity 62.5%; Pred. No. 5.8e-51;  
Matches 451; Conservative 0; Mismatches 256; Indels 15; Gaps 4;  
QY 1469 GCGATGGCGTATGGGTGTCATCCGCGCTGGACTCCACCCCTGAGGCTCTACCATGCACACA 1528  
Db 1 GGGATGGCGTGGGTGTCCTCAATCGCTTGGATTCTACGCTCCGCTCTATCATGCACACA 60  
QY 1529 CGCACCAGCATCTACAGGACGCTGGACATTGAGCCCTACGTACCAAGATGCTAGGCACTG 1588  
Db 61 CTTATCAACATCTACAGGATGTGGACATTGAGCCTTATGTAAGCAAAATGTTAGGTACTG 120  
QY 1589 GCAAGCTGGGTTCTCTTCTGACGCATCACGGCCCTGCTTGTGGGGGAGCCGGCTCT 1648  
Db 121 GAAACTGGGCTTCTCTTTTGTGAGAAATTACAGCTCTTATGGTGTCTTGTAAATCGTTGT 180  
QY 1649 GGGTGGGCACGGCAACGGAGTGGTCTATCTCCATCCCCCTGACAGAGACTGTGGTCTCTGC 1708  
Db 181 GGGTGGGGACAGGAAATGGTGTCTATTATCTCATCCCATGACAGAAACCGTAATCCTCC 240  
QY 1709 ACCGAGGCCAGCTCTCTGGGGTCCGAGGCCAATAAGACATCCCCACCTCTGGGGAGGGG 1768  
Db 241 ACCAGGGACGTTTACTGGGGTGGGGGCAATAAAA-----CCTCAGGTGTACAGGAA 294  
QY 1769 CCCGTCCCGGGGCATCATCCAGTGTATGGCGATGACAGAGTGACA---GGGCGGCCA 1825  
Db 295 ATCGTCTCTGAAAGTGAATCGGTGTATGTTGTATGATGAACAGTGATAAAGTGAATCCAG 354  
QY 1826 GCAGCTTCAATCCCTACTGTCTCCATGGCCAGGCCAGCTATGCTTCCATGGGCACCGCG 1885  
Db 355 GGACATTTATACCCCTATTGTTCAATGGCATGCACAGCTTGTCTTCCATGGGCACCGGG 414  
QY 1886 ATGCCGTGAAGTTCTTTGCTCGGTGCCAGGGAACGTGCTGGCCACCTGGAATGGCAGTG 1945  
Db 415 ATGCTGTGAATTTCTTTGTGGCAGTCCAGGTCAAGTCAATCAGCCCAAAAGTAGCAGTA 474  
QY 1946 TGCTGGACAGCCAGCCAGGGCCCTGGGCCAGTGTGCCCTGCTCGGAGGTGAGGGCC 2005  
Db 475 GTGGCAGGATCTGACGGGTGACAAAGCAGGGCCATCTGCACAGGAGCCTGGTAGTCAGA 534  
QY 2006 AGAAGCTGCGGAACGTGTGTGTGCTGACCGGCGGGAGGGCTACATCGACTTCCGCAATG 2065  
Db 535 CGCCCTTGAAGTCTATGCTTGTCTATCATGTGGAGGAGGGCTACATCGACTTCCGAATGG 594  
QY 2066 GAGACGAGAGGACGACGAGACGGAGGGCGCAGGGGACATGAGCCAGGTGAAGCCCG 2125  
Db 595 GTGATGAAG-----GTGGAGAATCAGAACTTCTTGGAGAGGATCTTCCACTTGAACCTTC 649  
QY 2126 TGCTGTCAAGGACGAGCGCAGTCATCATCTGTGTGGCAGGTGTCTTACACCCCGAGT 2185  
Db 650 TG-TCAACAAAGCAGAAAGGAGTCACTTGATAGTGTGGCAAGTGTATGGCAATGATG 708  
QY 2186 GA 2187  
Db 709 GA 710

RESULT 6  
US-10-085-783A-1790

; Sequence 1790, Application US/10085783A  
; Publication No. US20040037841A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liew, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2002  
; CURRENT APPLICATION NUMBER: US/10/085,783A  
; CURRENT FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1790  
; LENGTH: 310  
; TYPE: DNA  
; ORGANISM: Human  
US-10-085-783A-1790

Query Match 6.5%; Score 241.8; DB 13; Length 310;  
Best Local Similarity 94.0%; Pred. No. 6.8e-51;  
Matches 296; Conservative 0; Mismatches 12; Indels 7; Gaps 4;  
QY 2824 CTGGGGCAGCTGCTTGAGAACAGAGACTGCTACCCCATCTGCCCCATGCAGGAGGCTCT 2883  
Db 1 CTGGGGCAGCTGCTTGAGAACAGAGACTGCTACCCCATCTGCCCCATGCAGGAGGCTCT 60  
QY 2884 TGCCAGCCCGTCTTGACCCGTGTCCCCCAGGCTCTGCTGGGCGAGAGACTCACCTTG 2943  
Db 61 TGCCAGCCCGTCTTGACCCGTGTCCCCCAGGCTCTGCTGGGCGAGAGACTCACCTTG 120  
QY 2944 GAGAGTGGGCCCT-GGAGTCTGTCTCCCTCCCAAGAGCCCGCAGGGT-GGGATTCTCAG 3001  
Db 121 GAGAGTGGGCCCTCGGAGTGTCTCCCTCCCAAGAGCCCGCAGGGTGGGATTTCTCAG 180  
QY 3002 GCTGCCAGGGCAGGCCAGGCTCAGGAAGAGGGGAGGCCCTTGGCCTCTCCGGGATCA 3061  
Db 181 GCTGTAGTGCAGGCCAGGCTCTCAGGAGAGGGGAGG---CCCTGGCTCTCCGGGATCA 237  
QY 3062 GTCCTAGGACACAGGCTCAGCCTCAGGTTGATGGGGATGATGTCTCCCGGGCCTGCC 3121  
Db 238 GTCCTAGGACACAGGCTCAGCCTCAGGTTGATGGGGATGATGTCTCCCGGG---CTGCC 295  
QY 3122 TCCTGCACGGGGCTC 3136  
Db 296 TCCTGCACGGGGCTC 310

RESULT 7  
US-10-242-535A-1790  
; Sequence 1790, Application US/10242535A  
; Publication No. US20040013663A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liew, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2005  
; CURRENT APPLICATION NUMBER: US/10/242,535A  
; CURRENT FILING DATE: 2002-09-12  
; PRIOR APPLICATION NUMBER: US 10/085,783  
; PRIOR FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994  
; SOFTWARE: PatentIn version 3.2

```
; SEQ ID NO 1790
; LENGTH: 310
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-1790

Query Match      6.5%; Score 241.8; DB 16; Length 310;
Best Local Similarity 94.0%; Pred. No. 6.8e-51;
Matches 296; Conservative 0; Mismatches 12; Indels 7; Gaps 4;

QY 2824 CTGGGGCAGCTGCTTGAGAACAGAGACTGCTACCCCATCCTGCCCATGCAGGAGGCTCT 2883
Db 1 CTGGGGCAGCTGCTTGAGAACAGAGACTGCTACCCCATCCTGCCCATGCAGGAGGCTCT 60

QY 2884 TGCCAGCCCCGTTCTGACCCGTTGCCCCAGGCTCTGCCCTGGGCGAGAAGACTCACCTTG 2943
Db 61 TGCCAGCCCCGTTCTGACCCGTTGCCCCAGGCTCTGCCCTGGGCGAGAAGACTCACCTTG 120

QY 2944 GAGGAGTGGGCCCT-GGAGTCTCTGTCCTCCCTCCAGAACGCCCCAGGGT-GGGATTCTCAG 3001
Db 121 GAGGAGTGGGCCCTCGGAGTGCTGTCTCCCTCCAGAACGCCCCAGGGTGGGATTCTCAG 180

QY 3002 GCTGCCAGGGCAGGCCCTCAGGAAGAGGGAGGCCCTGGCCTCTCCGGGATCA 3061
Db 181 GCTGTCA GTGCAGGCCCTCAGGAGTCTCAGGAGAGGGAGG--CCCTGGCTCTCCGGGATCA 237

QY 3062 GTCTTAGGACACAGGCTCAGCCTCAGGTTGATGGGGATGATGTGCTCCCGGGCTGCC 3121
Db 238 GTCTTAGGACACAGGCTCAGCCTCAGGTTGATGGGGATGATGTGCTCCCGGG--CTGCC 295

QY 3122 TCCTGCACGGGGCTC 3136
Db 296 TCCTGCACGGGGCTC 310
```

```
RESULT 8
US-10-264-237-807
; Sequence 807, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PAL31P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; PRIOR FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 807
; LENGTH: 554
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (361)..(416)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (416)..(416)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-264-237-807

Query Match      6.1%; Score 226; DB 16; Length 554;
Best Local Similarity 72.1%; Pred. No. 7e-47;
Matches 373; Conservative 1; Mismatches 137; Indels 6; Gaps 6;

QY 1069 AGCAGTGTGCACCCACCATGTGGCTGGAGCCCCAGACGGCTGGCTCTATGTGCATCG 1128
Db 29 AGTAGTCTTTTACCAACTATGTGGCTTGAGAGCTCAAAATGGCTGTTTGTATGTCCATCA 88
```

```
QY 1129 GCTGTGGCCAACTGGAAGAGTGCCTGCACTCCATCAAGCTGAAGGATTCTGTGCTGAGC 1188
Db 89 TCTGTAGCCAGTGGAGGAATGTCTCCATTCCATTAAACTTAAAGATTCCGATTCTCAGT 148

QY 1189 CTGCTGCATGTCAAAGGCCGCTGTGCTGGTGGCTCTGCGGACGGACCCCTGGCCATCTTC 1248
Db 149 ATTGTACACGTGAAGGAATCGTGTAGTAGCCCTGGCTGACGGCACCCCTTGCAATCTTT 208

QY 1249 CACCGTGTGAAGATGGCCAGTGGGATCTGAGCAACTATCACCTAATGGACCTGGGCCAC 1308
Db 209 CACAGAGAGTGGATGGGCACTGGGATTTGTCAAACCTATCACCTCTTAGACCTTGGACGG 268

QY 1309 CCGCACCACTCCATCCGCTGCATGGCTGTTGTGTACGACCGCGTGTGGTGTGGCTACAAG 1368
Db 269 CCTCATCATTCATCCGTTGCTGACTGTGTGATGACATGACAAAGTCTGGTGTGGCTATAGG 328

QY 1369 AACAAAGTGCACGTCAATCCAGCCCAAGACCAT-GCAGATAGAGAAAGTCATTTGACGCCCA 1427
Db 329 ARCAAAATCTATGTGGTGCAGCCAAAGGCCAATNGAAATAGAGAAATCTTTTGATGCACA 388

QY 1428 CCCGCGGGGAGAGCCAGGTGCGGCAGCTGGCGTGGATCGCGATGGCGTATGGGTGTC 1487
Db 389 TCCCAGGAAGGAGAGCCAAAGTGGACA-NTTGGTGGTGGGGATGGCGTGTGGGTCTC 447

QY 1488 CATCCGCTTGGGATTTCTACGGTCCGCTCTCTATCATGGCACACACTTATTCAACATCTACA 1544
Db 448 CATTCGCTTGGGATTTCTACGGTCCGCTCTCTATCATGGCACACACTTATTCAACATCTACA 507

QY 1545 GGACGT-GGACATTGAGCCCTTACGTCAGCAAGATGCT 1580
Db 508 GGATTGGGACATTGAGCCTTATGTAAAGCAAAATGTT 544
```

```
RESULT 9
US-09-917-800A-965/c
; Sequence 965, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 965
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
```



```
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6852
; LENGTH: 189
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-040-862-6852

Query Match          5.1%; Score 189; DB 15; Length 189;
Best Local Similarity 100.0%; Pred. No. 1.3e-37;
Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2833 CTGCTTGAGAACAGAGACTGTACCCCATCTGCCCCATGCGGCGAGAAGACTCACCTTGGAGGAGTGG 2952
          |||||||
Db      1 CTGCTTGAGAACAGAGACTGTACCCCATCTGCCCCATGCGGCGAGAAGACTCACCTTGGAGGAGTGG 60

QY      2893 CGTTCTGACCCGTCGTCCTCCCGCAGGCTCTGCTGGGCGAGAAGACTCACCTTGGAGGAGTGG 2952
          |||||||
Db      61 CGTTCTGACCCGTCGTCCTCCCGCAGGCTCTGCTGGGCGAGAAGACTCACCTTGGAGGAGTGG 120

QY      2953 GCCCTGGAGTCCTGTCCCTCCCGCAGGCTCTGCTGGGCGAGAAGACTCACCTTGGAGGAGTGG 3012
          |||||||
Db      121 GCCCTGGAGTCCTGTCCCTCCCGCAGGCTCTGCTGGGCGAGAAGACTCACCTTGGAGGAGTGG 180

QY      3013 AGGCCCCAGG 3021
          |||||||
Db      181 AGGCCCCAGG 189

RESULT 12
US-10-057-475B-6852
; Sequence 6852, Application US/10057475B
; Publication No. US20040002068A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Clapper, Jonathan David
; APPLICANT: Wang, Aijun
; APPLICANT: Ordonez, Nadia
; APPLICANT: Carter, Lauren
; APPLICANT: McNeill, Patricia Dianne
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-014402US
; CURRENT APPLICATION NUMBER: US/10/057,475B
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
```

```
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 10979
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6852
; LENGTH: 189
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-057-475B-6852

Query Match          5.1%; Score 189; DB 16; Length 189;
Best Local Similarity 100.0%; Pred. No. 1.3e-37;
Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2833 CTGCTTGAGAACAGAGACTGTACCCCATCTGCCCCATGCGGCGAGAAGACTCACCTTGGAGGAGTGG 2892
          |||||||
Db      1 CTGCTTGAGAACAGAGACTGTACCCCATCTGCCCCATGCGGCGAGAAGACTCACCTTGGAGGAGTGG 60

QY      2893 CGTTCTGACCCGTCGTCCTCCCGCAGGCTCTGCTGGGCGAGAAGACTCACCTTGGAGGAGTGG 2952
          |||||||
Db      61 CGTTCTGACCCGTCGTCCTCCCGCAGGCTCTGCTGGGCGAGAAGACTCACCTTGGAGGAGTGG 120

QY      2953 GCCCTGGAGTCCTGTCCCTCCCGCAGGCTCTGCTGGGCGAGAAGACTCACCTTGGAGGAGTGG 3012
          |||||||
Db      121 GCCCTGGAGTCCTGTCCCTCCCGCAGGCTCTGCTGGGCGAGAAGACTCACCTTGGAGGAGTGG 180

QY      3013 AGGCCCCAGG 3021
          |||||||
Db      181 AGGCCCCAGG 189

RESULT 13
US-10-154-884B-6852
; Sequence 6852, Application US/10154884B
; Publication No. US20040005561A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc W.
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-013521US
; CURRENT APPLICATION NUMBER: US/10/154,884B
; CURRENT FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
```

```
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 11290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6852
; LENGTH: 189
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-154-884B-6852

Query Match          5.1%; Score 189; DB 16; Length 189;
Best Local Similarity 100.0%; Pred. No. 1.3e-37;
Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2833 CTGCTTGAGAACAGAGACTGCTACCCCATCTGCCCCATGCAGGAGGCTCTTGCCAGCCC 2892
Db 1 CTGCTTGAGAACAGAGACTGCTACCCCATCTGCCCCATGCAGGAGGCTCTTGCCAGCCC 60

QY 2893 CGTTCTGACCCGTGTCCCCCAGGCTCTGCCTGGGCAGAGACTCACCTTGGAGAGTGG 2952
Db 61 CGTTCTGACCCGTGTCCCCCAGGCTCTGCCTGGGCAGAGACTCACCTTGGAGAGTGG 120

QY 2953 GCCCTGGAGTCTGTCCCTCCAGAAAGCCCCCAGGGTGGGATTCTCAGGCTGCCAGGGC 3012
Db 121 GCCCTGGAGTCTGTCCCTCCAGAAAGCCCCCAGGGTGGGATTCTCAGGCTGCCAGGGC 180

QY 3013 AGGCCCCAGG 3021
Db 181 AGGCCCCAGG 189
```

```
RESULT 14
US-09-954-456-819
; Sequence 819, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 819
; LENGTH: 6289
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-819

Query Match          3.2%; Score 116.6; DB 9; Length 6289;
Best Local Similarity 52.0%; Pred. No. 3.3e-19;
Matches 325; Conservative 0; Mismatches 279; Indels 21; Gaps 2;

QY 1083 CACCATGTGGCTGGAGCCAGAACGGCTGGCTCTATGTGCACCTCGGCTGTGGCCAACTG 1142
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4012 CTCGTGTGGCTGGCACTGAGGATGGCTGTGCCACGTGTACCACTCTCCGACAGCAT 4071
QY 1143 GAAGAAAGTGCCTGCACCTCCATCAAGCTGAAGAT-----TCTGTGCTGAGCCTGGTGCA 1196
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4072 CCGTGACCGCAGGAACAGCATGAAGCTCCAGCATGCGGCCCTCTGTGACCTGCATCTTGTA 4131

QY 1197 TGTCAAAGGCCGTGTGTGGTGGCTCTGGGGAGGGGACCCCTGGCCATCTTCCACCGTGG 1256
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4132 TCTGAATAACCAAGTGTGTGTGTCTGTGGCAATGGAGAGCTTGTGGTCTACCAAGGGA 4191

QY 1257 TGAAGATGGCCAGTGGGATCTGAGCAACTATCACCTAATGGACCTGGGCCACCCCGCACCA 1316
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4192 AGCAGGCCATTTCTGGGACCCCGAGAACTTCAATCAGTGACCTTGGGCACCCAGGGGAG 4251

QY 1317 CTCATCCGCTGCATGGCTGTTGTGTACGACCGCTGTGGTGTGGCTACAGAACAAAGGT 1376
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4252 CCCCATCACCAAGATGTTATCTGTGGTGGCGGCTGTGGTGTGGCTGCCAGAACCGAGT 4311

QY 1377 GCACGTATCCAGCCCCAAGACCATGCAGATAGAGAAGTCATTTGACGCCCCACCCGCGGCG 1436
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4312 CTTGTCTTGACCCCTGACACGCTGCAGCTGGAGCACATGTTTACGTGGGTGAGGATTC 4371

QY 1437 GGAGAGCCAGGTGGGCAGCTGGCGTGGATCGCGGATGGCGGTATGGGTGTCCATCCGCT 1496
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4372 AAGCCGCTGCGTGGCTTGCATGGTGGACTCCAGCCTGGGTGTGGTGAATTGAAAGG 4431

QY 1497 GGACTCCACCTGAGGCTCTACCATGCACACAGCACCGACCATCTACAGGACGTGGACAT 1556
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4432 TAGTGCCCAAGTGTGTCTTACCATCCAGACACCTTTGAGCAGCTGGCAGAGTAGACGT 4491

QY 1557 TGAGCCCTACGTACGATCAGCAAGATGCTAGG-----CACTGGCAAGCTGGGTTT 1601
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4492 CACTCCTCCGTCGACAGGATGCTGGCAGGCTCGGATGCCATCATCCGGCAGCACAAAGC 4551

QY 1602 CTCCTTCGTACGATCAGCGCCCTGCTTGTGCGGGGACCGGCTCTGGTGGGCACCGG 1661
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4552 TGCCTGTCTGGAATCAGAGCTGCTGGTGTGTGAGGAGCTGTGTGGTGGGCACCGAG 4611

QY 1662 CAACGGAGTGGTCATCTCCATCCCC 1686
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4612 TGCTGGTGTGCTCCTCACCATGCCC 4636

RESULT 15
US-09-954-456-1216
; Sequence 1216, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
```

```
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1216
; LENGTH: 6289
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-954-456-1216

Query Match      3.2%; Score 116.6; DB 9; Length 6289;
Best Local Similarity 52.0%; Pred. No. 3.3e-19;
Matches 325; Conservative 0; Mismatches 279; Indels 21; Gaps 2;

QY 1083 CACCATGTGGCTGGAGAGCCAGAACGGCTGGCTTATGTGACATCGGCTGTGGCCAACTG 1142
Db 4012 CTCGGTGTGGCTGGGCACATGAGGATGGCTGTGTCCACGTGTACCAAGTCTCCGACAGCAT 4071

QY 1143 GAAGAAGTGCCTGCACTCCATCAAGCTGAAGGAT-----TCTGTGCTGAGCCTGGTGCA 1196
Db 4072 CCGTGACCGCAGGAACAGCATGAAGCTCCAGCATGCGGCCCTCTGTGACCTGCATCTTGTA 4131

QY 1197 TGTCAAAGGCCGTGTGCTGGTGTGGCTGTGGCGGACCGTGGCCATCTTCCACCGTGG 1256
Db 4132 TCTGAATAACCAAGGTGTTGTGTCTCTGGCCAAATGGAGAGCTTGTGCTTACCAAAGGGA 4191

QY 1257 TGAAGATGCCAGTGGGATCTGAGCAACTATCACCTAATGGACCTGGGCCACCCCGACCA 1316
Db 4192 AGCAGGCCATTTCTGGGACCCCCAGAACTTCAAATCAGTGACCTTGGGCACCCAGGGGAG 4251

QY 1317 CTCCATCCGCTGCATGGCTGTTGTGTAGCACCGCTGTGGTGTGGCTACAAGAACAGGT 1376
Db 4252 CCCCATCACCAAGATGGTATCTGTGGTGGCGGCTGTGGTGTGGCTGCCAGAACCGAGT 4311

QY 1377 GCACGTCAATCCAGCCCAAGACCATGCAGATAGAGAAGTCATTTGACGCCACCCCGCGCG 1436
Db 4312 CCTTGTCTGAGCCCTGACACGCTGCAGCTGGAGCACATGTTTTACGTGGGTCAAGGATTC 4371

QY 1437 GGAGAGCCAGGTGCGGACGTGGCTGGTGGATCGGCGATGGCGTATGGGTGTCCATCCGCCT 1496
Db 4372 AAGCCGCTGCGTGGCTTGATGGTGGACTCCAGCCTGGGTGTGGGTGACATTGAAAGG 4431

QY 1497 GGACTCCACCCCTGAGGCTCTACCATGCACACGCAACAGCATCTACAGGACGTGGACAT 1556
Db 4432 TAGTGCCCAACGTGTGTCTCTACCATCCAGACACCTTTGAGCAGCTGGCAGAAAGTAGCGT 4491

QY 1557 TGAGCCCTACGTACGCAAGATGCTAGG-----CACTGGCAAGCTGGGTTT 1601
Db 4492 CACTCCTCCGCTGCACAGGATGCTGGCAGGCTCGGATGCCATCATCCGGCAGCACAAAGC 4551

QY 1602 CTCCTTCGTACGCATCACGGCCCTGCTTGTGCGGGCAGCGGCTCTGGGTGGGCACCCGG 1661
Db 4552 TGCCTGTCTGCGAATCACAGGCTGTGTGTGTGAGGAGCTGTGGGTGGGCACCAG 4611

QY 1662 CAACGGAGTGTTCATCTCCATCCCC 1686
Db 4612 TGCTGGTGTGCTCCTCACCATGCCC 4636
```